Observation and measurement of Higgs boson decays to WW* with the ATLAS detector


DOI
10.1103/PhysRevD.92.012006

Publication date
2015

Document Version
Final published version

Published in
Physical Review D. Particles and Fields

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).
I. INTRODUCTION

In the Standard Model of particle physics (SM), the Higgs boson results from the Brout-Englert-Higgs mechanism [1] that breaks the electroweak symmetry [2] and gives mass to the W and Z gauge bosons [3]. It has a spin parity of $0^+$, with couplings to massive particles that are precisely determined by their measured masses. A new particle compatible with the spin and gauge-boson couplings of the SM Higgs boson was discovered in 2012 by the ATLAS and CMS experiments at the LHC using the $ZZ^*$, $\gamma\gamma$, and $WW^*$ final states [4–8]. Measurements of the particle’s mass [8,9] yield a value of approximately 125 GeV, consistent with the mass of the SM Higgs boson provided by a global fit to electroweak measurements [10]. Evidence for production of this boson at the Tevatron [11] and for its decay to fermions at the LHC [12] are also consistent with the properties of the SM Higgs boson.

The direct observation of the Higgs boson in individual decay channels provides an essential confirmation of the SM predictions. For a Higgs boson with a mass of 125 GeV, the $H \rightarrow WW^*$ decay has the second largest branching fraction (22%) and is a good candidate for observation. The sequential decay $H \rightarrow WW^* \rightarrow \ell^+\ell^-\nu\bar{\nu}$, where $\ell^+$ is an electron or muon, is a sensitive experimental signature. Searches for this decay produced the first direct limits on the mass of the Higgs boson at a hadron collider [13,14], and measurements following the boson discovery are among the most precise in determining its couplings and spin [5–7].

The dominant Higgs boson production mode in high-energy $pp$ collisions is gluon fusion (ggF), where the interacting gluons produce a Higgs boson predominantly through a top-quark loop. The next most abundant production mechanism, with a factor of 12 reduction in rate, is the fusion of vector bosons radiated by the interacting quarks into a Higgs boson (vector-boson fusion or VBF). At a further reduced rate, a Higgs boson can be produced in association with a W or Z boson (vector and Higgs boson production or VH). The leading-order production processes are depicted in Fig. 1.

This paper describes the observation and measurement of the Higgs boson in its decay to a pair of $W$ bosons, with the Higgs boson produced by the ggF and VBF processes at center-of-mass energies of 7 and 8 TeV. The ggF production process probes Higgs boson couplings to heavy quarks, while the VBF and VH processes probe its couplings to heavy quarks and bosons through the total width. To constrain these couplings, the rates of the ggF and VBF $H \rightarrow WW^*$ processes are measured—individually and combined—and normalized by the SM predictions for the ATLAS measured mass value of 125.36 GeV [9] to obtain the “signal strength” parameters $\mu_{ggF}$ and $\mu_{VBF}$. The total cross section for each process is also measured, along with fiducial cross sections for the ggF process.
A prior measurement of these processes with the same data set yielded a combined result of $\mu = 1.0 \pm 0.3$ [5]. The results presented here supersede this measurement and contain improvements in signal acceptance, background determination and rejection, and signal yield extraction. Together, these improvements increase the expected significance of an excess of $H \to WW^*$ decays over background from 3.7 to 5.8 standard deviations, and reduce the expected relative uncertainty on the corresponding $\mu$ measurement by 30%.

The paper is organized as follows. Section II provides an overview of the signal and backgrounds, and of the data analysis strategy. Section III describes the ATLAS detector and data, and the event reconstruction. The selection of events in the different final states is given in Sec. IV. Sections V and VI discuss the modeling of the signal and the background processes, respectively. The signal yield extraction and the various sources of systematic uncertainty are described in Sec. VII. Section VIII provides the event extraction and the likelihood fit. The definition of event categories based on lepton flavor and jet multiplicity, as illustrated in Fig. 2. In the final step of the analysis, a profile likelihood fit is simultaneously performed on all categories in order to extract the signal from the backgrounds and measure its yield.

The Drell-Yan (DY) process is the dominant source of events with two identified leptons, and contributes to the signal final state when there is a mismeasurement of the net particle momentum in the direction transverse to the beam (individual particle momentum in this direction is denoted $p_T$). The DY background is strongly reduced in events with different-flavor leptons ($ee/\mu\mu$), as these arise through fully leptonic decays of $\tau$-lepton pairs with a small branching fraction and reduced lepton momenta. The analysis thus separates $ee$ events from those with same-flavor leptons ($ee/\mu\mu$) in the event selection and the likelihood fit.

Pairs of top quarks are also a prolific source of lepton pairs, which are typically accompanied by high-momentum jets. Events are removed if they have a jet identified to contain a $b$-hadron decay ($b$-jet), but the $t\bar{t}$ background remains large due to inefficiencies in the $b$-jet identification algorithm. Events are therefore categorized by the number of jets. The top-quark background provides a small

---

**TABLE I.** Backgrounds to the $H \to WW^*$ measurement in the final state with two charged leptons ($\ell = e$ or $\mu$) and neutrinos, and no jet that contains a $b$-quark. Irreducible backgrounds have the same final state; other backgrounds are shown with the features that lead to this final state. Quarks from the first or second generation are denoted as $q$, and $j$ represents a jet of any flavor.

<table>
<thead>
<tr>
<th>Name</th>
<th>Process</th>
<th>Feature(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW</td>
<td>WW</td>
<td>Irreducible</td>
</tr>
<tr>
<td>Top quarks</td>
<td>$t\bar{t}$ ($t \to Wb\bar{b}$)</td>
<td>Unidentified $b$-quarks</td>
</tr>
<tr>
<td></td>
<td>$t$ ($t \to Wq\bar{q}$)</td>
<td>Unidentified $b$-quark, $q$ or $b$ misidentified as $\ell$; unidentified $b$-quarks</td>
</tr>
<tr>
<td>Misidentified leptons (Misid)</td>
<td>$Wj$ ($W + jet(s)$)</td>
<td>$j$ misidentified as $\ell$</td>
</tr>
<tr>
<td></td>
<td>$jj$ (Multijet production)</td>
<td>$jj$ misidentified as $\ell\ell$; misidentified neutrinos</td>
</tr>
<tr>
<td>Other dibosons</td>
<td>$VV$</td>
<td>$\gamma$ misidentified as $e$</td>
</tr>
<tr>
<td></td>
<td>$\gamma\gamma$</td>
<td>Unidentified lepton(s)</td>
</tr>
<tr>
<td></td>
<td>$W\gamma$, $WZ$, $ZZ \to \ell\ell\ell\ell$</td>
<td>Irreducible</td>
</tr>
<tr>
<td></td>
<td>$Z\gamma$</td>
<td>$\gamma$ misidentified as $e$; unidentified lepton</td>
</tr>
<tr>
<td>Drell-Yan (DY)</td>
<td>$ee/\mu\mu$</td>
<td>Misidentified neutrinos</td>
</tr>
<tr>
<td></td>
<td>$\tau\tau$</td>
<td>Irreducible</td>
</tr>
</tbody>
</table>

...
contribution to the zero-jet category but represents a significant fraction of the total background in categories with one or more jets.

In events with two or more jets, the sample is separated by signal production process (“VBF-enriched” and “ggF-enriched”). The VBF process is characterized by two quarks scattered at a small angle, leading to two well-separated jets with a large invariant mass [15]. These and other event properties are inputs to a boosted decision tree (BDT) algorithm [16] that yields a single-valued discriminant to isolate the VBF process. A separate analysis based on a sequence of individual selection criteria provides a cross-check of the BDT analysis. The ggF-enriched sample contains all events with two or more jets that do not pass either of the VBF selections.

Due to the large Drell-Yan and top-quark backgrounds in events with same-flavor leptons or with jets, the most sensitive signal region for ggF production is $n_j=0$ in $e\mu$, while for VBF production it is $n_j \geq 2$ in $e\mu$. These two samples are underlined. The $e\mu$ samples with $n_j \leq 1$ are further subdivided as described in the text.

In events with two or more jets, the sample is separated by signal production process (“VBF-enriched” and “ggF-enriched”). The VBF process is characterized by two quarks scattered at a small angle, leading to two well-separated jets with a large invariant mass [15]. These and other event properties are inputs to a boosted decision tree (BDT) algorithm [16] that yields a single-valued discriminant to isolate the VBF process. A separate analysis based on a sequence of individual selection criteria provides a cross-check of the BDT analysis. The ggF-enriched sample contains all events with two or more jets that do not pass either of the VBF selections.

Due to the large Drell-Yan and top-quark backgrounds in events with same-flavor leptons or with jets, the most sensitive signal region is in the $e\mu$ zero-jet final state. The dominant background to this category is WW production, which is effectively suppressed by exploiting the properties of W boson decays and the spin-0 nature of the Higgs boson (Fig. 3). This property generally leads to a lepton pair with a small opening angle [17] and a correspondingly low invariant mass $m_{\ell\ell}$, broadly distributed in the range below $m_W/2$. The dilepton invariant mass is used to select signal events, and the signal likelihood fit is performed in two ranges of $m_{\ell\ell}$ in $e\mu$ final states with $n_j \leq 1$.

Other background components are distinguished by $p_T^{\ell\ell}$, the magnitude of the transverse momentum of the lower-$p_T$ lepton in the event (the “subleading” lepton). In the signal process, one of the W bosons from the Higgs boson decay is off shell, resulting in relatively low subleading lepton $p_T$ (peaking near 22 GeV, half the difference between the Higgs and W boson masses). In the background from W bosons produced in association with a jet or photon (misreconstructed as a lepton) or an off-shell photon producing a low-mass lepton pair (where one lepton is not reconstructed), the $p_T^{\ell\ell}$ distribution falls rapidly with increasing $p_T$. The $e\mu$ sample is therefore subdivided into three regions of subleading lepton $p_T$ for $n_j \leq 1$. The jet and photon misidentification rates differ for electrons and muons, so this sample is further split by subleading lepton flavor.

Because of the neutrinos produced in the signal process, it is not possible to fully reconstruct the invariant mass of the final state. However, a “transverse mass” $m_T$ [18] can be calculated without the unknown longitudinal neutrino momenta:

$$m_T = \sqrt{E_T^{\ell\ell} + p_T^{\ell\ell}}^2 - |p_T^{\ell\ell} + p_T^{\nu\nu}|^2,$$

where $E_T^{\ell\ell} = \sqrt{(p_T^{\ell\ell})^2 + (m_{\ell\ell})^2}$, $p_T^{\nu\nu}$ ($p_T^{\ell\ell}$) is the vector sum of the neutrino (lepton) transverse momenta, and $p_T^{\ell\ell}$ ($p_T^{\nu\nu}$) is its modulus. The distribution has a kinematic upper bound at the Higgs boson mass, effectively separating Higgs boson production from the dominant nonresonant WW and top-quark backgrounds. For the VBF analysis, the transverse mass is one of the inputs to the BDT distribution used to fit for the signal yield. In the ggF and cross-check VBF analyses, the signal yield is obtained from a direct fit to the $m_T$ distribution for each category.

Most of the backgrounds are modeled using Monte Carlo samples normalized to data, and include theoretical uncertainties on the extrapolation from the normalization region.
to the signal region, and on the shape of the distribution used in the likelihood fit. For the $W + \text{jet(s)}$ and multijet backgrounds, the high rates and the uncertainties in modeling misidentified leptons motivate a model of the kinematic distributions based on data. For a few minor backgrounds, the process cross sections are taken from theoretical calculations. Details of the background modeling strategy are given in Sec. VI.

The analyses of the 7 and 8 TeV data sets are separate, but use common methods where possible; differences arise primarily because of the lower instantaneous and integrated luminosities in the 7 TeV data set. As an example, the categorization of 7 TeV data does not include a ggF-enriched category for events with at least two jets, since the expected significance of such a category is very low. Other differences are described in the text or in dedicated subsections.

III. DATA SAMPLES AND RECONSTRUCTION

This section begins with a description of the ATLAS detector, the criteria used to select events during data-taking (triggers) and the data sample used for this analysis. A description of the event reconstruction follows. The Monte Carlo simulation samples used in this analysis are described next, and then differences between the 2012 and 2011 analyses are summarized.

A. Detector and data samples

The ATLAS detector [19] is a multipurpose particle detector with approximately forward-backward symmetric cylindrical geometry. The experiment uses a right-handed coordinate system with the origin at the nominal $p\bar{p}$ interaction point at the center of the detector. The positive $x$ axis is defined by the direction from the origin to the center of the LHC ring, the positive $y$ axis points upwards, and the $z$ axis is along the beam direction. Cylindrical coordinates ($r, \phi$) are used in the plane transverse to the beam; $\phi$ is the azimuthal angle around the beam axis. Transverse components of vectors are not indicated by the subscript T. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$.

The inner tracking detector (ID) consists of a silicon-pixel detector, which is closest to the interaction point, a silicon-microstrip detector surrounding the pixel detector—both covering $|\eta| < 2.5$—and an outer transition-radiation straw-tube tracker (TRT) covering $|\eta| < 2$. The TRT provides substantial discriminating power between electrons and pions over a wide energy range. The ID is surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field.

A highly segmented lead/liquid-argon (LAr) sampling electromagnetic calorimeter measures the energy and the position of electromagnetic showers with $|\eta| < 3.2$. The LAr calorimeter includes a presampler (for $|\eta| < 1.8$) and three sampling layers, longitudinal in shower depth, up to $|\eta| < 2.5$. The LAr sampling calorimeters are also used to measure hadronic showers in the endcap ($1.5 < |\eta| < 3.2$) and electromagnetic and hadronic showers in the forward regions ($3.1 < |\eta| < 4.9$), while a steel/scintillator tile calorimeter measures hadronic showers in the central region ($|\eta| < 1.7$).

The muon spectrometer (MS) surrounds the calorimeters and is designed to detect muons in the pseudorapidity range $|\eta| < 2.7$. The MS consists of one barrel ($|\eta| < 1.05$) and two endcap regions. A system of three large superconducting air-core toroid magnets, each with eight coils, provides a magnetic field with a bending integral of about 2.5 T m in the barrel and up to 6 T m in the endcaps. Monitored drift tube chambers in both the barrel and endcap regions and cathode strip chambers covering $2.0 < |\eta| < 2.7$ are used as precision-measurement chambers, whereas resistive plate chambers in the barrel and thin gap chambers in the endcaps are used as trigger chambers, covering $|\eta| < 2.4$. The chambers are arranged in three layers, so high-$p_T$ particles traverse at least three stations with a lever arm of several meters.

A three-level trigger system selects events to be recorded for offline analysis. The first level (level-1 trigger) is hardware based, and the second two levels (high-level trigger) are software based. This analysis uses events selected by triggers that required either a single lepton or two leptons (dilepton). The single-lepton triggers had more restrictive lepton identification requirements and higher $p_T$ thresholds than the dilepton triggers. The specific triggers used for the 8 TeV data with the corresponding thresholds at the hardware and software levels are listed in Table II. Offline, two leptons—either $ee, \mu\mu$, or $e\mu$—with opposite charge are required. The leading lepton ($l_1$) is required to have $p_T \geq 22$ GeV and the subleading lepton ($l_2$) is required to have $p_T \geq 10$ GeV.

TABLE II. Summary of the minimum lepton $p_T$ trigger requirements (in GeV) during the 8 TeV data-taking. For single-electron triggers, the hardware and software thresholds are either 18 and 24i or 30 and 60, respectively. The “i” denotes an isolation requirement that is less restrictive than the isolation requirement imposed in the offline selection. For dilepton triggers, the pair of thresholds corresponds to the leading and subleading lepton, respectively; the “$\mu, \mu$” dilepton trigger requires only a single muon at level-1. The “and” and “or” are logical.

<table>
<thead>
<tr>
<th>Name</th>
<th>Level-1 trigger</th>
<th>High-level trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single lepton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e$</td>
<td>18 or 30</td>
<td>24i or 60</td>
</tr>
<tr>
<td>$\mu$</td>
<td>15</td>
<td>24i or 36</td>
</tr>
<tr>
<td>Dilepton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e, e$</td>
<td>10 and 10</td>
<td>12 and 12</td>
</tr>
<tr>
<td>$\mu, \mu$</td>
<td>15</td>
<td>18 and 8</td>
</tr>
<tr>
<td>$e, \mu$</td>
<td>10 and 6</td>
<td>12 and 8</td>
</tr>
</tbody>
</table>
The efficiency of the trigger selection is measured using a tag-and-probe method with a data sample of $Z/\gamma^* \rightarrow ee, \mu\mu$ candidates. For muons, the single-lepton trigger efficiency varies with $\eta$ and is approximately 70% for $|\eta| < 1.05$ and 90% for $|\eta| > 1.05$. For electrons, the single-lepton trigger efficiency increases with $p_T$, and its average is approximately 90%. These trigger efficiencies are for leptons that satisfy the analysis selection criteria described below. Dilepton triggers increase the signal acceptance by allowing lower leading-lepton $p_T$ thresholds to be applied offline while still remaining in the kinematic range that is in the plateau of the trigger efficiency. The trigger efficiencies for signal events satisfying the selection criteria described in Sec. IV are 95% for events with a leading electron and a subleading muon, 81% for events with a leading muon and subleading electron, 89% for $\mu\mu$ events, and 97% for ee events. These efficiencies are for the $n_j = 0$ category; the efficiencies are slightly larger for categories with higher jet multiplicity.

The data are subjected to quality requirements: events recorded when the relevant detector components were not operating correctly are rejected. The resulting integrated luminosity is 20.3 fb$^{-1}$ taken at $\sqrt{s} = 8$ TeV in 2012 and 4.5 fb$^{-1}$ at 7 TeV in 2011. The mean number of inelastic collisions per bunch crossing had an average value of 20 in 2012 and 9 in 2011. Overlapping signals in the detector due to multiple interactions—as well as signals due to interactions occurring in other nearby bunch crossings—are referred to as “pile-up.”

**B. Event reconstruction**

The primary vertex of each event must have at least three tracks with $p_T \geq 400$ MeV and is selected as the vertex with the largest value of $\Sigma(p_T)^2$, where the sum is over all the tracks associated with that particular vertex.

Muon candidates are identified by matching a reconstructed ID track with a reconstructed MS track [20]. The MS track is required to have a track segment in at least two layers of the MS. The ID tracks are required to have at least a minimum number of associated hits in each of the ID subdetectors to ensure good track reconstruction. This analysis uses muon candidates referred to as “combined muons” in Ref. [20], in which the track parameters of the MS track and the ID track are combined statistically. Muon candidates are required to have $|\eta| < 2.50$. The efficiencies for reconstructing and identifying combined muons are provided in Ref. [20].

Electron candidates are clusters of energy deposited in the electromagnetic calorimeter that are associated with ID tracks [21]. All candidate electron tracks are fitted using a Gaussian sum filter [22] (GSF) to account for bremsstrahlung energy losses. The GSF fit reduces the difference between the energy measured in the calorimeter and the momentum measured in the ID and improves the measured electron direction and impact parameter resolutions. The impact parameter is the lepton track’s distance of closest approach in the transverse plane to the reconstructed position of the primary vertex. The electron transverse energy is computed from the cluster energy and the track direction at the interaction point.

Electron identification is performed in the range $|\eta| < 2.47$, excluding the transition region between the barrel and endcap EM calorimeters, $1.37 < |\eta| < 1.52$. The identification is based on criteria that require the longitudinal and transverse shower profiles to be consistent with those expected for electromagnetic showers, the track and cluster positions to match in $\eta$ and $\phi$, and signals of transition radiation in the TRT. The electron identification has been improved relative to that described in Ref. [5] by adding a likelihood-based method in addition to the selection-based method. The likelihood allows the inclusion of discriminating variables that are difficult to use with explicit requirements without incurring significant efficiency losses. Detailed discussions of the likelihood identification and selection-based identification and the corresponding efficiency measurements can be found in Ref. [23]. Electrons with $10 < E_T < 25$ GeV must satisfy the “very tight” likelihood requirement, which reduces backgrounds from light-flavor jets and photon conversions by 35% relative to the selection-based identification with the same signal efficiency. For $E_T > 25$ GeV, where misidentification backgrounds are less important, electrons must satisfy the “medium” selection-based requirement. The single-lepton trigger applies the medium selection-based requirements. Using a likelihood-based selection criterion in addition to this selection-based requirement would result in a loss of signal efficiency without sufficient compensation in background rejection. Finally, additional requirements reduce the contribution of electrons from photon conversions by rejecting electron candidates that have an ID track that is part of a conversion vertex or that do not have a hit in the innermost layer of the pixel detector.

To further reduce backgrounds from misidentified leptons, additional requirements are imposed on the lepton impact parameter and isolation. The significance of the transverse impact parameter, defined as the measured transverse impact parameter $d_0$ divided by its estimated uncertainty $\sigma_{d_0}$, is required to satisfy $|d_0|/\sigma_{d_0} < 3.0$; the longitudinal impact parameter $z_0$ must satisfy the requirement $|z_0 \sin \theta| < 0.4$ mm for electrons and 1.0 mm for muons.

Lepton isolation is defined using track-based and calorimeter-based quantities. Details about the definition of electron isolation can be found in Ref. [23]. The track isolation is based on the scalar sum $\Sigma p_T$ of all tracks with $p_T > 400$ MeV for electrons ($p_T > 1$ GeV for muons) that are found in a cone in $\eta$-$\phi$ space around the lepton, excluding the lepton track. Tracks used in this scalar sum are required to be consistent with coming from the primary vertex. The cone size is $\Delta R = 0.4$ for leptons with
$p_T < 15$ GeV, where $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$, and $\Delta R = 0.3$ for $p_T > 15$ GeV. The track isolation selection criterion uses the ratio of the $\Sigma p_T$ divided by the electron $E_T$ (muon $p_T$). This ratio is required to be less than 0.06 for leptons with $10 < p_T < 15$ GeV, and this requirement increases monotonically to 0.10 for electrons (0.12 for muons) for $p_T > 25$ GeV.

The calorimeter isolation selection criterion—like the track isolation—is based on a ratio. The relative calorimetric isolation for electrons is computed as the sum of the cluster transverse energy $\Sigma E_T$ of surrounding energy deposits in the electromagnetic and hadronic calorimeters inside a cone of $\Delta R = 0.3$ around the candidate electron cluster, divided by the electron $E_T$. The cells within $0.125 \times 0.175$ in $\eta \times \phi$ around the electron cluster barycenter are excluded. The pile-up and underlying-event contributions to the calorimeter isolation are estimated and subtracted event by event. The electron relative calorimetric isolation upper bound varies monotonically with electron $E_T$: it is 0.20 for $10 < E_T < 15$ GeV, increasing to 0.28 for $E_T > 25$ GeV. In the case of muons, the relative calorimetric isolation discriminant is defined as the $\Sigma E_T$ calculated from calorimeter cells within $\Delta R = 0.3$ of the muon candidate, and with energy above a noise threshold, divided by the muon $p_T$. All calorimeter cells within the range $\Delta R < 0.05$ around the muon candidate are excluded from $\Sigma E_T$. A correction based on the number of reconstructed primary vertices in the event is made to $\Sigma E_T$ to compensate for extra energy due to pile-up. The muon relative calorimetric isolation upper bound also varies monotonically with muon $p_T$: it is 0.06 for $10 < p_T < 15$ GeV, increasing to 0.28 for $p_T > 25$ GeV.

The signal efficiencies of the impact parameter and isolation requirements are measured using a tag-and-probe method with a data sample of $Z/\gamma^* \rightarrow ee, \mu\mu$ candidates. The efficiencies of the combined impact parameter and isolation requirements range from 68% (60%) for electrons (muons) with $10 < p_T < 15$ GeV to greater than 90% (96%) for electrons (muons) with $p_T > 25$ GeV.

Jets are reconstructed using the anti-$k_t$ sequential recombination clustering algorithm [24] with a radius parameter $R = 0.4$. The inputs to the reconstruction are three-dimensional clusters of energy [25,26] in the calorimeter. The algorithm for this clustering suppresses noise by keeping only cells with a significant energy deposit and their neighboring cells. To take into account the differences in calorimeter response to electrons and photons and hadrons, each cluster is classified, prior to the jet reconstruction, as coming from an electromagnetic or hadronic shower using information from its shape. Based on this classification, the local cell signal weighting calibration method [27] applies dedicated corrections for the effects of calorimeter noncompensation, signal losses due to noise threshold effects, and energy lost in regions that are not instrumented. Jets are corrected for contributions from in-time and out-of-time pile-up [28], and the position of the primary interaction vertex. Subsequently, the jets are calibrated to the hadronic energy scale using $p_T$ and $\eta$-dependent correction factors determined in a first pass from simulation and then refined in a second pass from data [26,27]. The systematic uncertainties on these correction factors are determined from the same control samples in data.

To reduce the number of jet candidates originating from pile-up vertices, a requirement is imposed on the jet vertex fraction, denoted $JVF$: for jets with $p_T < 50$ GeV and $|\eta| < 2.4$, more than 50% of the summed scalar $p_T$ of tracks within $\Delta R = 0.4$ of the jet axis must be from tracks associated with the primary vertex ($JVF > 0.50$) [29]. No $JVF$ selection requirement is applied to jets that have no associated tracks.

For the purposes of classifying an event in terms of jet multiplicity $n_j$, a jet is required to have $p_T^j > 25$ GeV for $|\eta|^j < 2.4$, and $p_T^j > 30$ GeV if $2.4 \leq |\eta|^j < 4.5$. The increased threshold in the higher-$|\eta|$ region suppresses jets from pile-up. The two highest-$p_T$ jets ($j_1, j_2$, ordered in $p_T$) are the “VBF jets” used to compute dijet variables in the VBF-enhanced $n_j \geq 2$ category.

Additional jets not counted in $n_j$ have lower thresholds in three scenarios. First, those used to reject events because they lie in the $\eta$ range spanned by the two leading jets in the VBF-enriched selection (see Sec. IVC) are considered if they have $p_T^j > 20$ GeV. Second, the jets for $b$-jet identification—described below—are required to have $p_T^j > 20$ GeV and $|\eta|^j < 2.4$. Third, the jets used for the calculation of soft hadronic recoil (see Sec. IV A and the $f_{\text{recoil}}$ definition therein) are required to have $p_T^j > 10$ GeV and have no $JVF$ requirement. The calibration procedure described above is applied only to jets with $p_T^j > 20$ GeV. Jets with $10 \text{ GeV} < p_T^j < 20 \text{ GeV}$ are used only in the $f_{\text{recoil}}$ definition, and the efficiency for the requirements on this quantity are measured directly from the data, so the analysis is not sensitive to the modeling of the energy scale of these soft jets in the Monte Carlo simulation.

The identification of $b$-quark jets ($b$-jets) is limited to the acceptance of the ID ($|\eta| < 2.5$). The $b$-jets are identified with a multivariate technique—the MV1 algorithm [30]—that is based on quantities that separate $b$ and $c$ jets from “light jets” arising from light-flavor quarks and gluons. The inputs [31] to this algorithm use quantities such as the presence of secondary vertices, the impact parameters of tracks, and the topologies of weak heavy-quark decays. The efficiency for identifying $b$-jets is measured [32] in a large data sample of dilepton $t\bar{t}$ pair candidates. An operating point that is 85% efficient for identifying $b$-jets is adopted. At this operating point, the probability of misidentifying a light jet as a $b$-jet is 10.3%.

Two leptons or a lepton and a jet may be close in $\eta, \phi$ space. The following procedure is adopted in the case of
overlapping objects. Electron candidates that have tracks that extend to the MS are removed. If a muon candidate and an electron candidate are separated by $\Delta R < 0.1$, then the muon is retained, and the electron is removed. These cases usually indicate a muon that has undergone bremsstrahlung in the ID material or calorimeter. A high-$p_T$ electron is always also reconstructed as a jet, so if an electron and the nearest jet are separated by less than $\Delta R = 0.3$, the jet is removed. In contrast, if a muon and a jet are separated by less than $\Delta R = 0.3$, the muon candidate is removed, as it is more likely to be a nonprompt muon from heavy-flavor decay. Finally, due to early bremsstrahlung, a prompt electron may produce more than one electron candidate in its vicinity. In the case of two electrons separated by less than $\Delta R = 0.1$, the electron candidate with larger $E_T$ is retained.

The signature of a high-momentum neutrino is a momentum imbalance in the transverse plane. The reconstruction of this “missing” transverse momentum [33] is calculated as the negative vector sum of the momentum of objects selected according to ATLAS identification algorithms, such as leptons, photons, and jets, and of the remaining “soft” objects that typically have low values of $p_T$. The calculation can thus be summarized as

$$E_T^{\text{miss}} = -\left( \sum_{\text{selected}} p_T + \sum_{\text{soft}} p_T \right), \quad (2)$$

where the reconstruction of soft objects and the choice of selected objects differ between different methods of evaluating the missing transverse momentum. Three methods of reconstruction are used in this analysis; $E_T^{\text{miss}}$ is used to represent one particular method, as described below.

The large coverage in rapidity ($y$) of the calorimeter and its sensitivity to neutral particles motivate a calorimeter-based reconstruction of the missing transverse momentum. Selected objects are defined as the leptons selected by the analysis, and photons and jets with $E_T > 20$ GeV. The transverse momenta of these objects are added vectorially using object-specific calibrations. For the remaining soft objects, calibrated calorimeter cluster energy measurements are used to determine their net transverse momentum. The resulting missing transverse momentum is denoted $E_T^{\text{miss}}$.

The significant pile-up present in the data degrades the resolution of the calorimeter-based measurement of missing transverse momentum. An $O(20\%)$ improvement in resolution is obtained using a track-based measurement of the soft objects, where the tracks are required to have $p_T > 0.5$ GeV and originate from the primary vertex. Tracks associated with identified leptons or jets are not included, as these selected objects are added separately to the calculation of the missing transverse momentum. This reconstruction of missing transverse momentum, denoted $p_T^{\text{miss}}$, is used in the final fit to the $m_T$ distribution and improves the signal resolution relative to the $E_T^{\text{miss}}$ used for the previous measurement [5]. Figure 4 shows the simulated resolution for the magnitude of $E_T^{\text{miss}}$ and $p_T^{\text{miss}}$ ($E_T^{\text{miss}}$ and $p_T^{\text{miss}}$ respectively), and for $m_T$ in the $n_j = 0$ category, all evaluated by subtracting the reconstructed quantity from the corresponding quantity obtained using generated leptons and neutrinos in ggF $H \rightarrow WW^{*}$ events. The rms of the $m_T$ difference decreases from 19 to 14 GeV when using $p_T^{\text{miss}}$ instead of $E_T^{\text{miss}}$ in the reconstruction. The improved resolution significantly increases the discrimination between signal and certain background processes (such as $W\gamma$).

A simplified version of $p_T^{\text{miss}}$ is used to suppress the Drell-Yan background in events with same-flavor leptons. This definition, denoted $p_T^{\text{miss\,(trk)}}$, differs from $p_T^{\text{miss}}$ in that

![MC sample for ggF $H \rightarrow WW^{*}$](image)

FIG. 4 (color online). Simulated resolutions of (a) missing transverse momentum and (b) $m_T$ for the ggF signal MC in the $n_j = 0$ category. The comparisons are made between the calorimeter-based reconstruction ($E_T^{\text{miss}}$) and the track-based reconstruction ($p_T^{\text{miss}}$) of the soft objects [see Eq. (2)]. The resolution is measured as the difference of the reconstructed (Reco) and generated (Gen) quantities; the rms values of the distributions are given with the legends in units of GeV.
the tracks associated with jets are also used, replacing the calorimeter-based jet measurement. This tends to align $p_T^{\text{miss (tk)}}$ with the jet(s) in Drell-Yan events, while in signal events $p_T^{\text{miss}}$ generally remains in the direction of the neutrinos. Incorporating the direction of $p_T^{\text{miss (tk)}}$ relative to the jet directions in the event selection thus improves Drell-Yan rejection.

The direction of $E_T^{\text{miss}}$ relative to the lepton and jet directions is also used to reject Drell-Yan, particularly the case of $\tau\tau$ production where $E_T^{\text{miss}}$ tends to align with a final-state lepton. A relative quantity $E_T^{\text{miss rel}}$ is defined as follows:

$$E_T^{\text{miss rel}} = \begin{cases} E_T^{\text{miss}} \sin \Delta \phi_{\text{near}} & \text{if } \Delta \phi_{\text{near}} < \pi/2, \\ E_T^{\text{miss}} & \text{otherwise}, \end{cases}$$

where $\Delta \phi_{\text{near}}$ is the azimuthal separation of the $E_T^{\text{miss}}$ and the nearest high-$p_T$ jet or lepton. A similar calculation defines $p_T^{\text{miss rel}}$ and $p_T^{\text{miss (tk)}}$.

C. Monte Carlo samples

Given the large number of background contributions to the signal region and the broadly peaking signal $m_T$ distribution, Monte Carlo modeling is an important aspect of the analysis. Dedicated samples are generated to evaluate all but the $W + j$ and multijet backgrounds, which are estimated using data (see Sec. VI C). Most samples use the POWHEG [34] generator to include corrections at next-to-leading order (NLO) in $\alpha_s$. In cases where higher parton multiplicities are important, ALPGEN [35] or SHERPA [36] provide merged calculations at tree level for up to five additional partons. In a few cases, only leading-order generators (such as ACERMC [37] or g2vV [38]) are available. Table III shows the generator and cross section used for each process.

The matrix-element-level Monte Carlo calculations are matched to a model of the parton shower, underlying event and hadronization, using either PYTHIA6 [39], PYTHIA8 [40], HERWIG [41] (with the underlying event modeled by JIMMY [42]), or SHERPA. Input parton distribution functions (PDFs) are taken from CT10 [43] for the POWHEG and SHERPA samples and CTEQ6L1 [44] for ALPGEN + HERWIG and ACERMC samples. The $Z/\gamma^*$ sample is reweighted to the MRSTmcsal PDF set [45].

Pile-up interactions are modeled with PYTHIA8, and the ATLAS detector response is simulated [46] using either GEANT4 [47] or GEANT4 combined with a parametrized GEANT4-based calorimeter simulation [48]. Events are filtered during generation where necessary, allowing up to $2\,ab^{-1}$ of equivalent luminosity for high cross section processes such as $Z/\gamma^*$ in the VBF category. The $ggF$ and VBF production modes for the $H \to WW^*$ signal are modeled with POWHEG + PYTHIA8 [49,50] at $m_T = 125$ GeV, and the corresponding cross sections are shown in Table III. A detailed description of these processes and their modeling uncertainties is given in Sec. V. The smaller contribution from the VH process, with subsequent $H \to WW^*$ decay, is also shown in Table III. Not shown are the $H \to \tau\tau$ MC samples, which have an even smaller contribution but are included in the signal modeling for completeness using the same generators as for the $H \to WW^*$ decay. The $H \to ZZ^*$ decay contributes negligibly after event selection and is not included in the analysis.

<table>
<thead>
<tr>
<th>Process</th>
<th>MC generator</th>
<th>$\sigma \cdot B$ (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ggF , H \to WW^*$</td>
<td>POWHEG + PYTHIA8</td>
<td>0.435</td>
</tr>
<tr>
<td>$VBF , H \to WW^*$</td>
<td>POWHEG + PYTHIA8</td>
<td>0.0356</td>
</tr>
<tr>
<td>$VH , H \to WW^*$</td>
<td>PYTHIA8</td>
<td>0.0253</td>
</tr>
<tr>
<td>$WW$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q\bar{q} \to WW$ and $qg \to WW$</td>
<td>POWHEG + PYTHIA6</td>
<td>5.68</td>
</tr>
<tr>
<td>$gg \to WW$</td>
<td>GG2VV + HERWIG</td>
<td>0.196</td>
</tr>
<tr>
<td>$(q\bar{q} \to W) + (q\bar{q} \to W)$</td>
<td>PYTHIA8</td>
<td>0.480</td>
</tr>
<tr>
<td>$q\bar{q} \to WW$</td>
<td>SHERPA</td>
<td>5.68</td>
</tr>
<tr>
<td>VBS $WW + 2$ jets</td>
<td>SHERPA</td>
<td>0.0397</td>
</tr>
<tr>
<td>Top quarks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>POWHEG + PYTHIA6</td>
<td>26.6</td>
</tr>
<tr>
<td>$Wt$</td>
<td>POWHEG + PYTHIA6</td>
<td>2.35</td>
</tr>
<tr>
<td>$t\bar{b}$</td>
<td>ACERMC + PYTHIA6</td>
<td>28.4</td>
</tr>
<tr>
<td>$tb$</td>
<td>POWHEG + PYTHIA6</td>
<td>1.82</td>
</tr>
<tr>
<td>Other dibosons (VV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W\gamma$ ($p_T^\gamma &gt; 8$ GeV)</td>
<td>ALPGEN + HERWIG</td>
<td>369</td>
</tr>
<tr>
<td>$W\gamma$ ($m_{\ell\ell} \leq 7$ GeV)</td>
<td>SHERPA</td>
<td>12.2</td>
</tr>
<tr>
<td>$WZ$ ($m_{\ell\ell} &gt; 7$ GeV)</td>
<td>POWHEG + PYTHIA8</td>
<td>12.7</td>
</tr>
<tr>
<td>VBS $WZ + 2$ jets</td>
<td>SHERPA</td>
<td>0.0126</td>
</tr>
<tr>
<td>($m_{\ell\ell} &gt; 7$ GeV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z\gamma$ ($p_T^\gamma &gt; 8$ GeV)</td>
<td>SHERPA</td>
<td>163</td>
</tr>
<tr>
<td>$Z\gamma$ (min $m_{\ell\ell} \leq 4$ GeV)</td>
<td>SHERPA</td>
<td>7.31</td>
</tr>
<tr>
<td>ZZ ($m_{\ell\ell} &gt; 4$ GeV)</td>
<td>POWHEG + PYTHIA8</td>
<td>0.733</td>
</tr>
<tr>
<td>ZZ $\to \ell\ell\nu\nu$ ($m_{\ell\ell} &gt; 4$ GeV)</td>
<td>POWHEG + PYTHIA8</td>
<td>0.504</td>
</tr>
<tr>
<td>Drell-Yan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z$ ($m_{\ell\ell} &gt; 10$ GeV)</td>
<td>ALPGEN + HERWIG</td>
<td>16500</td>
</tr>
<tr>
<td>VBF $Z + 2$ jets</td>
<td>SHERPA</td>
<td>5.36</td>
</tr>
<tr>
<td>($m_{\ell\ell} &gt; 7$ GeV)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cross sections are calculated for the dominant diboson and top-quark processes as follows: the inclusive $WW$ cross section is calculated to NLO in $\alpha_s$ with MCFM [51]; nonresonant gluon fusion is calculated and modeled to leading order (LO) in $\alpha_s$ with gg2VV, including both $WW$ and $ZZ$ production and their interference; $t\bar{t}$ production is normalized to the calculation at next-to-next-to-leading order (NNLO) in $\alpha_s$ with resummed of higher-order terms to the next-to-next-to-leading logarithms (NNLL), evaluated with TOP++2.0 [52]; and single-top processes are normalized to NNLL following the calculations from Refs. [53–55] for the $s$-channel, $t$-channel, and $Wt$ processes, respectively. The $t\bar{t}$, $Wt$, and single-top $s$-channel kinematics are modeled with POWHEG + PYTHIA6 [56–58], while ACERMC [37] is used for the single-top $t$-channel process. The $WW$ kinematics are modeled using the POWHEG + PYTHIA6 [59] sample for the $n_j \leq 1$ categories and the merged multileg SHERPA sample for the $n_j \geq 2$ categories. Section VI A describes this modeling and the normalization of the double parton interaction process ($q\bar{q} \to W) + (q\bar{q} \to W)$, which is modeled using the PYTHIA8 generator. For $WW$, $WZ$, and $ZZ$ production via nonresonant vector-boson scattering, the SHERPA generator provides the LO cross section and is used for event modeling. The negligible vector-boson scattering $ZZ$ process is not shown in the table but is included in the background modeling for completeness.

The process $W\gamma^*$ is defined as associated $W + Z/\gamma^*$ production, where there is an opposite-charge same-flavor lepton pair with invariant mass $m_{\ell\ell}$ less than 7 GeV. This process is modeled using SHERPA with up to one additional parton. The range $m_{\ell\ell} > 7$ GeV is simulated with POWHEG + PYTHIA8 [59] and normalized to the POWHEG cross section. The use of SHERPA for $W\gamma^*$ is due to the inability of POWHEG + PYTHIA8 to model invariant masses down to the dielectron production threshold. The SHERPA sample requires two leptons with $p_T > 5$ GeV and $|\eta| < 3$. The jet multiplicity is corrected using a SHERPA sample generated with $0.5 < m_{\ell\ell} < 7$ GeV and up to two additional partons, while the total cross section is corrected using the ratio of the MCFM NLO to SHERPA LO calculations in the same restricted mass range. A similar procedure is used to model $Z\gamma^*$, defined as $Z/\gamma^*$ pair production with one same-flavor opposite-charge lepton pair having $m_{\ell\ell} \leq 4$ GeV and the other having $m_{\ell\ell} > 4$ GeV.

The $W\gamma$ and $DY$ processes are modeled using ALPGEN + HERWIG with merged tree-level calculations of up to five jets. The merged samples are normalized to the NLO calculation of MCFM (for $W\gamma$) or the NNLO calculation of DYNNLO [60] (for $Z/\gamma^*$). The $W\gamma$ sample is generated with the requirements $p_T^{\gamma} > 8$ GeV and $\Delta R(\gamma, \ell) > 0.25$. A $W\gamma$ calculation at NNLO [61] finds a correction of less than 8% in the modeled phase space, which falls within the uncertainty of the NLO calculation.

A SHERPA sample is used to accurately model the $Z(\to \ell\ell)\gamma$ background. The photon is required to have $p_T^\gamma > 8$ GeV and $\Delta R(\gamma, \ell) > 0.1$; the lepton pair must satisfy $m_{\ell\ell} > 10$ GeV. The cross section is normalized to NLO using MCFM. Events are removed from the ALPGEN + HERWIG DY samples if they overlap with the kinematics defining the SHERPA $Z(\to \ell\ell)\gamma$ sample.

The uncertainties are discussed for each specific background in Sec. VI, and their treatment in the likelihood fit is summarized in Sec. VII.
with requirements on the missing transverse momentum distributions. Figure 5(a) shows the $E_{T,\text{rel}}^{\text{miss}}$ distribution in the $n_j \leq 1$ $ee/\mu\mu$ sample, where the dominant $Z/\gamma^* \to e\mu$ contribution is suppressed by the $E_{T,\text{rel}}^{\text{miss}} > 40$ GeV requirement. In the $n_j \leq 1$ and $n_j \geq 2$ ggF-enriched $ee$ samples, a $p_T^{\text{miss}} > 20$ GeV selection is applied to significantly reduce the $Z/\gamma^* \to \tau\tau$ background and the multijet backgrounds with misidentified leptons [see Figs. 5(b) and 5(c) for the $n_j \leq 1$ categories]. The $n_j \geq 2$ VBF-enriched $ee$ sample requires no missing transverse momentum selection, and thus recovers signal acceptance for the statistically limited VBF measurement. In the $ee/\mu\mu$ sample, more stringent selections are applied: $E_{T,\text{rel}}^{\text{miss}} > 45$ GeV and $p_T^{\text{miss}} > 40$ GeV. Table IV lists these so-called preselection criteria.

The different background composition as a function of jet multiplicity motivates the division of the data sample into the various $n_j$ categories. Figures 6(a) and 6(b) show the jet multiplicity distributions in the $ee/\mu\mu$ and $e\mu$ samples, respectively. The $Z/\gamma^* \to e\mu\mu$ background dominates the $n_j \leq 1$ $ee/\mu\mu$ samples even after the above-mentioned missing transverse momentum requirements. The top-quark background becomes more significant at higher jet multiplicities. Its suppression is primarily based on the $b$-jet multiplicity; the distribution is shown in Fig. 6(c) for the $e\mu$ sample.

In each of the $n_j$ and lepton-flavor categories, further criteria are applied to increase the precision of the signal measurement. Sections IV A to IV D present the discriminating distributions and the resulting event yields. The selections are also listed in Table IV along with the preselection. Section IV E details the selection modifications for the 7 TeV data analysis. Section IV F concludes with the distributions after all requirements are applied.

In this section, the background processes are normalized using control regions (see Sec. VI). The distributions in the figures and the rates in the tables for the signal process are normalized to the figures and the rates in the tables for the signal process using control regions (see Sec. VI).

A. $n_j = 0$ category

Events with a significant mismeasurement of the missing transverse momentum are suppressed by requiring $p_T^{\text{miss}}$ to point away from the dilepton transverse momentum ($\Delta\phi_{\ell\ell,\text{MET}} > \pi/2$). In the absence of a reconstructed jet to balance the dilepton system, the magnitude of the dilepton momentum $p_T^{\ell\ell}$ is expected to be small in DY events. A requirement of $p_T^{\ell\ell} > 30$ GeV further reduces the DY contribution while retaining the majority of the signal events, as shown for the $e\mu$ sample in Fig. 7(a). At this stage, the DY background is sufficiently reduced in the $e\mu$ sample, but still dominates in the $ee/\mu\mu$ one. In this latter sample, a requirement of $p_T^{\text{miss,\text{rel}}} > 40$ GeV is applied to provide further rejection against DY events.

The continuum $WW$ production and the resonant Higgs boson production processes can be separated by exploiting the spin-0 property of the Higgs boson, when combined with the $V-A$ nature of the $W$ boson decay, leads to a small opening angle between the charged leptons (see Sec. II). A requirement of $\Delta\phi_{\ell\ell} < 1.8$ reduces both the $WW$ and DY backgrounds while retaining 90% of the signal. A related requirement of $m_{\ell\ell} < 55$ GeV combines the small lepton opening angle with the kinematics of a low-mass Higgs boson ($m_H = 125$ GeV). The $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ distributions are shown for the $e\mu$ sample in Figs. 7(b) and 7(c), respectively.

An additional discriminant, $f_{\text{recose}}$, based on soft jets, is defined to reduce the remaining DY contribution in the $ee/\mu\mu$ sample. This residual DY background satisfies the event selection primarily when the measurement of the energy associated with partons from initial-state
TABLE IV. Event selection summary. Selection requirements specific to the $e\mu$ and $ee/\mu\mu$ lepton-flavor samples are noted as such (otherwise, they apply to both); a dash (-) indicates no selection. For the $n_j \geq 2$ VBF-enriched category, MET denotes all types of missing transverse momentum observables. Values are given for the analysis of 8 TeV data for $m_H = 125$ GeV; the modifications for 7 TeV are given in Sec. IV E. All energy-related values are in GeV.

<table>
<thead>
<tr>
<th>Objective</th>
<th>ggF-enriched</th>
<th>ggF-enriched</th>
<th>VBF-enriched</th>
<th>VBF-enriched</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_j = 0$</td>
<td>$n_j = 1$</td>
<td>$n_j \geq 2$</td>
<td>$n_j \geq 2$</td>
<td></td>
</tr>
<tr>
<td>Preselection</td>
<td>$p_T^{\ell_1} &gt; 22$ for the leading lepton $\ell_1$</td>
<td>$p_T^{\ell_2} &gt; 10$ for the subleading lepton $\ell_2$</td>
<td>$p_T^{\ell} &gt; 20$ for $e\mu$</td>
<td>No MET requirement for $e\mu$</td>
</tr>
<tr>
<td></td>
<td>Opposite-charge leptons</td>
<td></td>
<td>$E_T^{\text{miss}} &gt; 40$ for ee/\mu\mu</td>
<td>$p_T^{\text{miss}} &gt; 40$ for ee/\mu\mu</td>
</tr>
<tr>
<td>Reject backgrounds</td>
<td>$p_T^{\text{miss (trk)}} &gt; 40$ for ee/\mu\mu</td>
<td>$p_T^{\text{miss (trk)}} &gt; 35$ for ee/\mu\mu</td>
<td>-</td>
<td>$p_T^{\text{miss (trk)}} &gt; 40$ for ee/\mu\mu</td>
</tr>
<tr>
<td>DY</td>
<td>$f_{\text{recoil}} &lt; 0.1$ for ee/\mu\mu</td>
<td>$f_{\text{recoil}} &lt; 0.1$ for ee/\mu\mu</td>
<td>-</td>
<td>$E_T^{\text{miss}} &gt; 45$ for ee/\mu\mu</td>
</tr>
<tr>
<td></td>
<td>$\Delta\phi_{\ell\ell,\text{MET}} &gt; \pi/2$</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Misid</td>
<td>-</td>
<td>$m_T^{\ell} &gt; 50$ for $e\mu$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Top</td>
<td>$n_j = 0$</td>
<td>$n_b = 0$</td>
<td>$n_b = 0$</td>
<td>$n_b = 0$</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$p_T^{\text{sum}}$ inputs to BDT</td>
</tr>
<tr>
<td>VBF topology</td>
<td></td>
<td></td>
<td></td>
<td>- $\Sigma m_{\ell\ell}$ inputs to BDT</td>
</tr>
<tr>
<td>$H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$</td>
<td>$m_{\ell\ell} &lt; 55$</td>
<td>$m_{\ell\ell} &lt; 55$</td>
<td>$m_{\ell\ell} &lt; 55$</td>
<td>$m_{\ell\ell}$ inputs to BDT</td>
</tr>
<tr>
<td></td>
<td>$\Delta\phi_{\ell\ell} &lt; 1.8$</td>
<td>$\Delta\phi_{\ell\ell} &lt; 1.8$</td>
<td>$\Delta\phi_{\ell\ell} &lt; 1.8$</td>
<td>$\Delta\phi_{\ell\ell}$ inputs to BDT</td>
</tr>
<tr>
<td></td>
<td>No $m_T$ requirement</td>
<td>No $m_T$ requirement</td>
<td>No $m_T$ requirement</td>
<td>$m_T$ inputs to BDT</td>
</tr>
</tbody>
</table>
radiation is underestimated, resulting in an apparent imbalance of transverse momentum in the event. To further suppress such mismeasured DY events, jets with $p_T > 10$ GeV, within a $\pi/2$ wedge in $\phi$ (noted as $\wedge$) centered on $-p_T^{ll}$, are used to define a fractional jet recoil relative to the dilepton transverse momentum:

$$f_{\text{recoil}} = \frac{\sum_{\text{jets \ } j \in \wedge} jVF_j \cdot p_T^j}{p_T^{ll}}. \quad (4)$$

The jet transverse momenta are weighted by their associated $jVF$ value to suppress the contribution from jets originating from pile-up interactions. Jets with no associated tracks are assigned a weight of 1. The $f_{\text{recoil}}$ distribution is shown in Fig. 7(d); a requirement of $f_{\text{recoil}} < 0.1$ reduces the residual DY background in the $ee/\mu\mu$ sample by a factor of 7.

The expected signal and background yields at each stage of selection are shown in Table V, together with the observed yields. At the final stage, the table also shows the event yields in the range $\frac{3}{2} m_H < m_T < m_H$ where most of the signal resides. This $m_T$ selection is not used to extract the final results, but nicely illustrates the expected signal-to-background ratios in the different categories.

**B. $n_j = 1$ category**

The one-jet requirement significantly increases the top-quark background. Since top quarks decay to $Wb$, events with jets with $p_T > 20$ GeV are rejected if they are identified as containing a $b$-quark [$n_b = 0$, see Fig. 6(c)]. After this requirement, the $WW$ and the DY background processes are dominant in the sample, as shown in Table VI.

In the case of the $e\mu$ sample, a requirement is applied to the transverse mass defined for a single lepton $\ell_i$:

$$m_T^{\ell_i} = \sqrt{2p_T^{\ell_i} \cdot p_T^{\text{miss}} \cdot (1 - \cos \Delta \phi)}, \quad (5)$$

where $\Delta \phi$ is the angle between the lepton transverse momentum and $p_T^{\text{miss}}$. This quantity tends to have small values for the DY background and large values for the signal process. It also has small values for multijet production, where misidentified leptons are frequently measured with energy lower than the jets from which they originate. The $m_T^{\ell_i}$ distribution, chosen to be the larger of $m_T^{\ell_1}$ or $m_T^{\ell_2}$, is presented in Fig. 8(a), and shows a clear difference in shape between the DY and multijet backgrounds, which lie mostly at low values of $m_T^{\ell_i}$, and the other background processes. Thus, both the DY and multijet processes are substantially reduced with a requirement of $m_T > 50$ GeV in the $e\mu$ sample.

The requirement of a jet allows for improved rejection of the $Z/\gamma^* \rightarrow \tau\tau$ background. Using the direction of the measured missing transverse momentum, the mass of the $\tau$-lepton pair can be reconstructed using the so-called collinear approximation [63]. A requirement of $m_{\tau\tau} < m_Z - 25$ GeV significantly reduces the remaining DY contribution in the $e\mu$ sample, as can be seen in Fig. 8(b).

The remaining selection criteria ($p_{T,\text{rel}}^{\text{miss(ink)}}$, $f_{\text{recoil}}$, $m_{\ell\ell}$, $\Delta \phi_{\ell\ell}$) are the same as in the $n_j = 0$ category, except that $p_{T,\text{rel}}^{\ell\ell}$ is replaced with the magnitude of $p_{T,\text{rel}}^{\ell\ell} = p_T^{\ell\ell} + p_T^{ll}$ in the calculation of $f_{\text{recoil}}$, and the $p_{T,\text{rel}}^{\text{miss(ink)}}$ threshold is reduced to 35 GeV. The $m_{\ell\ell}$ and $\Delta \phi_{\ell\ell}$ distributions are shown in Figs. 8(c) and 8(d), respectively. Differences between the shapes of the signal or WW processes and the $Z/\gamma^*$ background processes are more apparent in the $\Delta \phi_{\ell\ell}$ distribution of the $e\mu + ee/\mu\mu$ events than of the $e\mu$ events.

**C. VBF-enriched $n_j \geq 2$ category**

The $n_j \geq 2$ sample contains signal events produced by both the VBF and ggF production mechanisms. This section focuses on the former; the next section focuses on the latter.

The sample is analyzed using a boosted decision tree multivariate method [16] that considers VBF Higgs boson production as signal and the rest of the processes as background, including ggF Higgs boson production. A cross-check analysis is performed using sequential selections on some of the variables that are used as inputs to the BDT. Table VII shows the sample composition after each of the selection requirements in the cross-check analysis. For the $WW$ and $Z/\gamma^* \rightarrow \tau\tau$ backgrounds, the table separates contributions from events with jets from QCD vertices and...
electroweak events with VBS or VBF interactions (see Table III).

The VBF process is characterized by the kinematics of the pair of tag jets ($j_1$ and $j_2$) and the activity in the rapidity gap between them. In general, this process results in two highly energetic forward jets with $\Delta y_{jj} > 3$, where $\Delta y_{jj} = |y_{j1} - y_{j2}|$. The invariant mass of this tag-jet pair combines $\Delta y_{jj}$ with $p_T$ information since $m_{jj} \approx \sqrt{p_T^1 \cdot p_T^2 e^{\Delta y_{jj}/2}}$ for large values of $\Delta y_{jj}$. Both $\Delta y_{jj}$ and $m_{jj}$ are input variables to the BDT; for the cross-check analysis, $\Delta y_{jj} > 3.6$ and $m_{jj} > 600$ GeV are required [see Figs. 9(a) and 9(b)].

The $\Delta y_{jj}$ gap defines a “central region,” where a relatively low level of hadronic activity is expected because the mediating weak bosons do not exchange color. The number of extra jets ($n_{\text{extra}}$) in the $\Delta y_{jj}$ gap quantifies the activity. Requiring the absence of such jets in this region is known as a “central-jet veto” [64] and it suppresses processes where the jets are produced via QCD radiation. The central-jet veto uses jets with $p_T > 20$ GeV, and this requirement is applied in both the BDT and cross-check analyses. The selection can be expressed in terms of jet centrality, defined as

$$C_{j3} = \left| \eta_{j3} - \frac{\Sigma \eta_{jj}}{2} \right| / \frac{\Delta \eta_{jj}}{2}. \quad (6)$$
TABLE V. Event selection for the $n_j = 0$ category in the 8 TeV data analysis. The selection is presented separately for the $e\mu$ and $ee/\mu\mu$ samples. The summary columns give the observed yields ($N_{\text{obs}}$), the expected background yields ($N_{\text{bkg}}$), their ratios, and the expected signal yields ($N_{\text{sig}}$). For the dominant backgrounds, the expected yields are normalized using control regions, as described in Sec. VI. The $N_{\text{sig}}$ values are given for $m_H = 125$ GeV and are subdivided into the $N_{ggF}$ and $N_{VBF}$ contributions. The composition columns give the contributions to $N_{\text{bkg}}$ (see Sec. VI). The requirements are imposed sequentially from top to bottom; entries are shown as 0.0 (-) if they are less than 0.1 (0.01) events. The entries are rounded to a precision commensurate with the statistical uncertainties due to the random error associated with the central value of the yield ($\text{stat}_{\text{obs}} = \sqrt{N_{\text{obs}}}$) and the sampling error associated with the finite sample size used for the prediction for background type $k$ ($\text{stat}_{\text{bkg,k}}$). The errors on $N_{\text{obs}}/N_{\text{bkg}}$ are due to the combined statistical uncertainty on $\text{stat}_{\text{obs}}$ and $\text{stat}_{\text{bkg}}$. Energy-related quantities are in GeV.

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{\text{obs}}/N_{\text{bkg}}$</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{bkg}}$</th>
<th>$N_{\text{sig}}$</th>
<th>$N_{\text{VBF}}$</th>
<th>$N_{\text{top}}$</th>
<th>$N_{\text{w}}$</th>
<th>$N_{\text{DY}}$</th>
<th>$N_{\text{DY}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$N_{\text{obs}}$</td>
<td>$N_{\text{bkg}}$</td>
<td>$N_{\text{VBF}}$</td>
<td>$N_{\text{DY}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi_{ee,\text{MET}} &gt; \pi/2$</td>
<td>1.01 ± 0.01</td>
<td>16423</td>
<td>16330</td>
<td>290</td>
<td>12.1</td>
<td>7110</td>
<td>820</td>
<td>407</td>
<td>1330</td>
</tr>
<tr>
<td>$p_T^j &gt; 30$</td>
<td>1.00 ± 0.01</td>
<td>16339</td>
<td>16270</td>
<td>290</td>
<td>12.1</td>
<td>7110</td>
<td>812</td>
<td>405</td>
<td>1330</td>
</tr>
<tr>
<td>$m_{ee} &lt; 55$</td>
<td>1.11 ± 0.02</td>
<td>3411</td>
<td>3060</td>
<td>224</td>
<td>6.3</td>
<td>1670</td>
<td>141</td>
<td>79</td>
<td>427</td>
</tr>
<tr>
<td>$\Delta \phi_{ee} &lt; 1.8$</td>
<td>1.12 ± 0.02</td>
<td>2642</td>
<td>2350</td>
<td>203</td>
<td>5.9</td>
<td>1500</td>
<td>132</td>
<td>75</td>
<td>278</td>
</tr>
<tr>
<td>$\frac{1}{4} m_H &lt; m_T &lt; m_H$</td>
<td>1.20 ± 0.04</td>
<td>1129</td>
<td>940</td>
<td>131</td>
<td>2.2</td>
<td>660</td>
<td>40</td>
<td>21</td>
<td>133</td>
</tr>
<tr>
<td>$ee/\mu\mu$ sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi_{ee,\text{MET}} &gt; \pi/2$</td>
<td>1.04 ± 0.01</td>
<td>38040</td>
<td>36520</td>
<td>163</td>
<td>7.2</td>
<td>3260</td>
<td>418</td>
<td>211</td>
<td>504</td>
</tr>
<tr>
<td>$p_T^j &gt; 30$</td>
<td>1.05 ± 0.01</td>
<td>35445</td>
<td>33890</td>
<td>163</td>
<td>7.1</td>
<td>3250</td>
<td>416</td>
<td>211</td>
<td>493</td>
</tr>
<tr>
<td>$m_{ee} &lt; 55$</td>
<td>1.06 ± 0.01</td>
<td>11660</td>
<td>11040</td>
<td>154</td>
<td>6.8</td>
<td>3010</td>
<td>394</td>
<td>201</td>
<td>396</td>
</tr>
<tr>
<td>$p_T^{\text{miss}} &gt; 40$</td>
<td>1.02 ± 0.02</td>
<td>2197</td>
<td>2160</td>
<td>117</td>
<td>4.3</td>
<td>1097</td>
<td>99</td>
<td>59</td>
<td>133</td>
</tr>
<tr>
<td>$\Delta \phi_{ee} &lt; 1.8$</td>
<td>1.01 ± 0.02</td>
<td>2127</td>
<td>2100</td>
<td>113</td>
<td>4.2</td>
<td>1068</td>
<td>96</td>
<td>57</td>
<td>122</td>
</tr>
<tr>
<td>$f_{\text{recoil}} &lt; 0.1$</td>
<td>1.01 ± 0.03</td>
<td>1108</td>
<td>1096</td>
<td>72</td>
<td>2.7</td>
<td>786</td>
<td>41</td>
<td>31</td>
<td>79</td>
</tr>
<tr>
<td>$\frac{1}{4} m_H &lt; m_T &lt; m_H$</td>
<td>0.99 ± 0.05</td>
<td>510</td>
<td>517</td>
<td>57</td>
<td>1.3</td>
<td>349</td>
<td>11</td>
<td>8</td>
<td>53</td>
</tr>
</tbody>
</table>

TABLE VI. Event selection for the $n_j = 1$ category in the 8 TeV data analysis (see Table V for presentation details).

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{\text{obs}}/N_{\text{bkg}}$</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{bkg}}$</th>
<th>$N_{\text{VBF}}$</th>
<th>$N_{\text{DY}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$N_{\text{obs}}$</td>
<td>$N_{\text{bkg}}$</td>
<td>$N_{\text{VBF}}$</td>
<td>$N_{\text{DY}}$</td>
</tr>
<tr>
<td>$e\mu$ sample</td>
<td>1.00 ± 0.01</td>
<td>20607</td>
<td>20700</td>
<td>131</td>
<td>32</td>
</tr>
<tr>
<td>$n_b = 0$</td>
<td>1.01 ± 0.01</td>
<td>10859</td>
<td>10790</td>
<td>114</td>
<td>26</td>
</tr>
<tr>
<td>$m_T^j &gt; 50$</td>
<td>1.01 ± 0.01</td>
<td>7368</td>
<td>7280</td>
<td>103</td>
<td>23</td>
</tr>
<tr>
<td>$m_{\tau} &lt; m_2 - 25$</td>
<td>1.02 ± 0.02</td>
<td>4574</td>
<td>4490</td>
<td>96</td>
<td>20</td>
</tr>
<tr>
<td>$m_{ee} &lt; 55$</td>
<td>1.05 ± 0.02</td>
<td>1656</td>
<td>1570</td>
<td>84</td>
<td>15</td>
</tr>
<tr>
<td>$\Delta \phi_{ee} &lt; 1.8$</td>
<td>1.10 ± 0.03</td>
<td>1129</td>
<td>1030</td>
<td>74</td>
<td>18</td>
</tr>
<tr>
<td>$\frac{1}{4} m_H &lt; m_T &lt; m_H$</td>
<td>1.21 ± 0.06</td>
<td>407</td>
<td>335</td>
<td>42</td>
<td>6.6</td>
</tr>
<tr>
<td>$ee/\mu\mu$ sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_b = 0$</td>
<td>1.05 ± 0.01</td>
<td>153444</td>
<td>14640</td>
<td>61</td>
<td>15</td>
</tr>
<tr>
<td>$m_T^j &gt; 50$</td>
<td>1.08 ± 0.02</td>
<td>9897</td>
<td>9140</td>
<td>53</td>
<td>12.1</td>
</tr>
<tr>
<td>$m_{\tau} &lt; 55$</td>
<td>1.16 ± 0.02</td>
<td>5127</td>
<td>4410</td>
<td>48</td>
<td>9.4</td>
</tr>
<tr>
<td>$p_T^{\text{miss}} &gt; 35$</td>
<td>1.14 ± 0.04</td>
<td>960</td>
<td>842</td>
<td>36</td>
<td>6.9</td>
</tr>
<tr>
<td>$\Delta \phi_{ee} &lt; 1.8$</td>
<td>1.14 ± 0.04</td>
<td>889</td>
<td>783</td>
<td>32</td>
<td>6.3</td>
</tr>
<tr>
<td>$f_{\text{recoil}} &lt; 0.1$</td>
<td>1.16 ± 0.05</td>
<td>467</td>
<td>404</td>
<td>20</td>
<td>3.6</td>
</tr>
<tr>
<td>$\frac{1}{4} m_H &lt; m_T &lt; m_H$</td>
<td>1.11 ± 0.10</td>
<td>143</td>
<td>129</td>
<td>14</td>
<td>2.0</td>
</tr>
</tbody>
</table>
A requirement of $C_l < 1$ is applied to each lepton in the BDT and cross-check analyses. The sum of lepton centralities $\Sigma C_l = C_{l1} + C_{l2}$ is used as an input to the BDT. The $C_{l1}$ distribution is shown in Fig. 9(c).

Top-quark pair production has a large cross section and the same final state as VBF Higgs boson production, with the exception that its jets result from $b$-quarks. A requirement of $n_b = 0$ with $p_T > 20$ GeV is made in the BDT and cross-check analyses. This requirement is made on all jets in the event regardless of classification as tag jets. Significant top-quark background still remains because of the limited $\eta$ coverage of the tracker, the $p_T$ threshold applied to the $b$-jets, and the inefficiency of the $b$-jet identification algorithm within the tracking region. Further reductions are achieved through targeted kinematic selections and the BDT.

The pair production of top quarks occurs dominantly through gluon-gluon annihilation, and is frequently accompanied by QCD radiation. This radiation is used as a signature to further suppress top-quark backgrounds using the summed vector $p_T$ of the final-state objects, $p_T^{\text{sum}} = p_T^{\ell\ell} + p_T^{\text{miss}} + \Sigma p_T^j$ where the last term is a sum of the transverse momenta of all jets in the event. Its magnitude $p_T^{\text{sum}}$ is used as input to the BDT and is required to be $p_T^{\text{sum}} < 15$ GeV in the cross-check analysis.

The sum of the four combinations of lepton-jet invariant mass, $\Sigma m_{\ell j} = m_{\ell 1,j1} + m_{\ell 1,j2} + m_{\ell 2,j1} + m_{\ell 2,j2}$, is also
used as an input to the BDT. In the VBF topology, tag jets are more forward whereas the leptons tend to be more central. This results in differences in the shapes of the $\Sigma m_{\ell j}$ distributions for the VBF signal and the background processes, as can be seen in Fig. 9(d). This variable is not used in the cross-check analysis.

The other BDT input variables are those related to the $H \rightarrow WW^* \rightarrow \ell^+\ell^-\nu\bar{\nu}$ decay topology ($m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$, $m_T$), which are also used in the $n_j \leq 1$ categories. The cross-check analysis requires $\Delta\phi_{\ell\ell} < 1.8$ and $m_{\ell\ell} < 50$ GeV.

Distributions from eight variables are input to the BDT: $\Sigma C_{\ell}$, $\Delta y_{jj}$, and $m_{jj}$ for VBF selection; $p_T^{\text{sum}}$ and $\Sigma m_{\ell j}$ for $t\bar{t}$ rejection; and $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, and $m_T$ for their sensitivity to the $H \rightarrow WW^* \rightarrow \ell^+\ell^-\nu\bar{\nu}$ decay topology. The BDT is trained after the common preselection criteria (as listed in Table IV) and the $n_b = 0$ requirement. This event selection stage corresponds to the $n_b = 0$ stage presented for the cross-check analysis in Table VII. Additional criteria, common to the BDT and cross-check analyses, are applied before the classification of the events based on the BDT output (described below). They include requirements on $m_{\ell\ell}$, $C_{\ell3}$, and $C_\ell$. The observed and expected event yields after all these requirements are shown in Table VIII(a) separately for the $ee$ and $e\mu/\mu\mu$ samples. The dominant background processes include $t\bar{t}$ and $Z/\gamma^* \rightarrow \ell^+\ell^- \nu\bar{\nu}$ production. The normalization factors, described in Sec. VI, are not applied to these backgrounds at this stage.

The BDT is trained using the MC samples after the above-mentioned selections. The training starts with a single decision tree where an event is given a score of $\pm1$ if it satisfies particular sets of decisions ($+1$ leaf contains signal-like events and $-1$ background-like ones).

A thousand such trees are built and in each iteration the weight of miscategorized events is relatively increased, or “boosted.” The final discriminant $O_{\text{BDT}}$ for a given event is the weighted average of the binary scores from the individual trees. The bin widths for the likelihood fit are optimized for the expected significance while keeping each bin sufficiently populated. The chosen configuration is four bins with boundaries at $[-1, -0.48, 0.3, 0.78, 1]$, and with corresponding bin numbers from 0 to 3. The lowest bin contains the majority of background events and has a very small signal-to-background ratio. It is therefore not used in the likelihood fit. The expected and observed event yields after the classification in bins of $O_{\text{BDT}}$ are shown in Table VIII(b). Here the background yields are scaled by their corresponding normalization factors.

Appendix B documents the BDT analysis in more detail. It includes the distributions of all the input variables (showing the comparison between the VBF signal and the background processes) used in the training of the BDT as well as the data-to-MC comparison after the BDT classification. The distributions are in agreement and show that the correlations are well understood.

### D. ggF-enriched $n_j \geq 2$ category

The sample of $n_j \geq 2$ events, which are neither in the VBF-enriched category for the BDT analysis nor in the cross-check analysis, are used to measure ggF production. In this category only the $ee$ final state is analyzed due to the relatively low expected significance in the $e\mu/\mu\mu$ sample. Table IX shows the signal and background yields after each selection requirement.

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{\text{obs}}/N_{\text{bkg}}$</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{bkg}}$</th>
<th>$N_{\text{signal}}$</th>
<th>$N_{\text{QCD}}$</th>
<th>$N_{\text{EW}}$</th>
<th>$N_{\text{top}}$</th>
<th>$N_{\text{QCD}}$</th>
<th>$N_{\text{EW}}$</th>
<th>$N_{\text{top}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$ sample</td>
<td>$1.00 \pm 0.00$</td>
<td>61434</td>
<td>61180</td>
<td>85</td>
<td>32</td>
<td>26</td>
<td>1350</td>
<td>68</td>
<td>51810</td>
<td>2970</td>
</tr>
<tr>
<td>$n_b = 0$</td>
<td>$1.02 \pm 0.01$</td>
<td>7818</td>
<td>7700</td>
<td>36</td>
<td>26</td>
<td>13</td>
<td>993</td>
<td>43</td>
<td>3000</td>
<td>367</td>
</tr>
<tr>
<td>$p_T^{\text{sum}} &lt; 15$</td>
<td>$1.03 \pm 0.01$</td>
<td>5787</td>
<td>5630</td>
<td>46</td>
<td>23</td>
<td>13</td>
<td>781</td>
<td>38</td>
<td>1910</td>
<td>270</td>
</tr>
<tr>
<td>$m_{\ell\ell} &lt; m_{jj} - 25$</td>
<td>$1.05 \pm 0.02$</td>
<td>3129</td>
<td>2970</td>
<td>40</td>
<td>20</td>
<td>9.9</td>
<td>484</td>
<td>22</td>
<td>1270</td>
<td>177</td>
</tr>
<tr>
<td>$m_{jj} &gt; 600$</td>
<td>$1.31 \pm 0.12$</td>
<td>131</td>
<td>100</td>
<td>100</td>
<td>2.3</td>
<td>8.2</td>
<td>18</td>
<td>8.9</td>
<td>40</td>
<td>5.3</td>
</tr>
<tr>
<td>$\Delta y_{jj} &gt; 3.6$</td>
<td>$1.33 \pm 0.13$</td>
<td>107</td>
<td>80</td>
<td>2.1</td>
<td>7.9</td>
<td>-</td>
<td>11.7</td>
<td>6.9</td>
<td>35</td>
<td>5.0</td>
</tr>
<tr>
<td>$C_{\ell3} &gt; 1$</td>
<td>$1.36 \pm 0.18$</td>
<td>58</td>
<td>43</td>
<td>1.3</td>
<td>6.6</td>
<td>-</td>
<td>6.9</td>
<td>5.6</td>
<td>14</td>
<td>3.0</td>
</tr>
<tr>
<td>$m_{\ell\ell}, \Delta\phi_{\ell\ell}, m_T$</td>
<td>$2.53 \pm 0.71$</td>
<td>14</td>
<td>5.5</td>
<td>0.8</td>
<td>4.7</td>
<td>-</td>
<td>1.0</td>
<td>0.5</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>$ee/\mu\mu$ sample</td>
<td>$0.99 \pm 0.01$</td>
<td>26949</td>
<td>27190</td>
<td>31</td>
<td>14</td>
<td>10.1</td>
<td>594</td>
<td>37</td>
<td>23440</td>
<td>1320</td>
</tr>
<tr>
<td>$n_b, p_T^{\text{sum}}, m_{\ell\ell}$</td>
<td>$1.03 \pm 0.03$</td>
<td>1344</td>
<td>1310</td>
<td>13</td>
<td>8.0</td>
<td>4.0</td>
<td>229</td>
<td>12.0</td>
<td>633</td>
<td>86</td>
</tr>
<tr>
<td>$m_{jj}, \Delta y_{jj}, C_{\ell3}, C_\ell$</td>
<td>$1.39 \pm 0.28$</td>
<td>26</td>
<td>19</td>
<td>0.4</td>
<td>2.9</td>
<td>0.0</td>
<td>3.1</td>
<td>3.1</td>
<td>5.5</td>
<td>1.0</td>
</tr>
<tr>
<td>$m_{\ell\ell}, \Delta\phi_{\ell\ell}, m_T$</td>
<td>$1.63 \pm 0.69$</td>
<td>6</td>
<td>3.7</td>
<td>0.3</td>
<td>2.2</td>
<td>0.0</td>
<td>0.4</td>
<td>0.2</td>
<td>0.6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Summary Composition of $N_{\text{bkg}}$
The initial selection, \( n_b = 0 \) and \( m_{tZ} < 25 \text{ GeV} \), is common to the other categories and reduces the top-quark and DY backgrounds. The ggF-enriched sample is forced to be mutually exclusive to the VBF-enriched sample by inverting at least one of the VBF-specific requirements: \( C_{j3} > 1 \), \( C_\ell < 1 \), or \( O_{\text{BDT}} > -0.48 \). A similar inversion is done for the cross-check analysis: \( \Delta y_{jj} > 3.6 \), \( m_{jj} > 600 \text{ GeV} \), \( n_{\text{extra-j}} = 0 \), or \( C_\ell < 1 \). Both sets of orthogonality requirements for the BDT and the cross-check are imposed for the \( n_j \geq 2 \) ggF-enriched category.

The resulting sample contains events in a region sensitive to VH production where the associated \( W \) or \( Z \) boson decays hadronically. This region is suppressed by rejecting events in the region of \( \Delta y_{jj} \leq 1.2 \) and \( |m_{jj} - 85| < 15 \text{ GeV} \).

Figure 10 shows the \( m_{\ell\ell} \) distribution after the VH orthogonality requirement. The \( H \to WW^* \to \ell\nu\ell\nu \) topological selections, \( m_{\ell\ell} < 55 \text{ GeV} \) and \( \Delta\phi_{\ell\ell} < 1.8 \), further reduce the dominant top-quark background by 70\%, resulting in a signal purity of 3.3\%.

### E. Modifications for 7 TeV data

The 7 TeV data analysis closely follows the selection used in the 8 TeV analysis. The majority of the differences can be found in the object definitions and identifications, as described in Sec. III B. The lower average pile-up allows...
TABLE VIII. Event selection for the $n_j \geq 2$ VBF-enriched category in the 8 TeV BDT data analysis (see Table V for presentation details). The event yields in (a) are shown after the preselection and the additional requirements applied before the BDT classification (see text). The event yields in (b) are given in bins in $O_{BDT}$ after the classification, the normalization factors are applied to the yields (see Table XX). In the specific case of (a), the normalization factors described in Sec. VI are not applied to the relevant backgrounds. The $N_{ggF}$, $N_{VBF}$, and $N_{VH}$ expected yields are shown separately.

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{obs}/N_{bkg}$</th>
<th>$N_{obs}$</th>
<th>$N_{bkg}$</th>
<th>$N_{ggF}$</th>
<th>$N_{VBF}$</th>
<th>$N_{VH}$</th>
<th>$N_{W^{WW}}$</th>
<th>$N_{W^{EW}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{w}}$</th>
<th>$N_{V^{ee}}$</th>
<th>$N_{Drell-Yan}$</th>
<th>$N_{DY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Before the BDT classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ee/\mu\mu$ sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e\mu$ sample</td>
<td>1.04 ± 0.04</td>
<td>718</td>
<td>689</td>
<td>13</td>
<td>15</td>
<td>2.0</td>
<td>90</td>
<td>11</td>
<td>327</td>
<td>42</td>
<td>29</td>
<td>23</td>
<td>31</td>
<td>2.2</td>
</tr>
<tr>
<td>ee sample</td>
<td>1.18 ± 0.08</td>
<td>469</td>
<td>397</td>
<td>6.0</td>
<td>7.7</td>
<td>0.9</td>
<td>37</td>
<td>3</td>
<td>132</td>
<td>17</td>
<td>5.2</td>
<td>1.2</td>
<td>10.1</td>
<td>168</td>
</tr>
</tbody>
</table>

(b) Bins in $O_{BDT}$

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{obs}/N_{bkg}$</th>
<th>$N_{obs}$</th>
<th>$N_{bkg}$</th>
<th>$N_{ggF}$</th>
<th>$N_{VBF}$</th>
<th>$N_{VH}$</th>
<th>$N_{W^{WW}}$</th>
<th>$N_{W^{EW}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{w}}$</th>
<th>$N_{V^{ee}}$</th>
<th>$N_{Drell-Yan}$</th>
<th>$N_{DY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ee sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bin 0 (not used)</td>
<td>1.02 ± 0.04</td>
<td>661</td>
<td>650</td>
<td>8.8</td>
<td>3.0</td>
<td>1.9</td>
<td>83</td>
<td>9</td>
<td>313</td>
<td>40</td>
<td>26</td>
<td>21</td>
<td>28</td>
<td>2.2</td>
</tr>
<tr>
<td>Bin 1</td>
<td>0.99 ± 0.16</td>
<td>37</td>
<td>37</td>
<td>3.0</td>
<td>4.2</td>
<td>0.1</td>
<td>5.0</td>
<td>1.0</td>
<td>17</td>
<td>3.1</td>
<td>3.3</td>
<td>1.8</td>
<td>2.6</td>
<td>-</td>
</tr>
<tr>
<td>Bin 2</td>
<td>2.26 ± 0.63</td>
<td>14</td>
<td>6.2</td>
<td>1.2</td>
<td>4.2</td>
<td>-</td>
<td>1.5</td>
<td>0.5</td>
<td>1.8</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>Bin 3</td>
<td>5.41 ± 2.32</td>
<td>6</td>
<td>1.1</td>
<td>0.4</td>
<td>3.1</td>
<td>-</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>ee/\mu\mu sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bin 0 (not used)</td>
<td>1.91 ± 0.08</td>
<td>396</td>
<td>345</td>
<td>3.8</td>
<td>1.3</td>
<td>0.8</td>
<td>33</td>
<td>2</td>
<td>123</td>
<td>16</td>
<td>4.1</td>
<td>1.1</td>
<td>8.8</td>
<td>137</td>
</tr>
<tr>
<td>Bin 1</td>
<td>0.82 ± 0.14</td>
<td>53</td>
<td>45</td>
<td>1.5</td>
<td>2.2</td>
<td>0.1</td>
<td>3.0</td>
<td>0.5</td>
<td>10.4</td>
<td>1.8</td>
<td>0.8</td>
<td>0.2</td>
<td>9.9</td>
<td>26</td>
</tr>
<tr>
<td>Bin 2</td>
<td>1.77 ± 0.49</td>
<td>14</td>
<td>7.9</td>
<td>0.6</td>
<td>2.5</td>
<td>-</td>
<td>0.8</td>
<td>0.3</td>
<td>1.1</td>
<td>0.2</td>
<td>0.2</td>
<td>-</td>
<td>0.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Bin 3</td>
<td>6.52 ± 2.87</td>
<td>6</td>
<td>0.9</td>
<td>0.2</td>
<td>1.7</td>
<td>-</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7</td>
<td>-</td>
</tr>
</tbody>
</table>

In the $n_j \geq 2$ category, only the VBF-enriched analysis is considered; it follows an approach similar to the 8 TeV version. It exploits the BDT multivariate method and it uses the same BDT classification and output binning as the 8 TeV data analysis. In the $e\mu$ sample, a two-bin fit of the $O_{BDT}$ is used (bins 2 and 3 are merged). In the $ee/\mu\mu$ sample, a one-bin fit is used (bins 1–3 are merged) due to the smaller sample size.

The background estimation, signal modeling, final observed and expected event yields, and the statistical analysis and results, are presented in the next sections.

**F. Summary**

This section described the event selection in the $n_j$ and lepton-flavor categories. Each of these is treated

TABLE IX. Event selection for the $n_j \geq 2$ ggF-enriched category in the 8 TeV data analysis (see Table V for presentation details). The $N_{ggF}$, $N_{VBF}$, and $N_{VH}$ expected yields are shown separately. The “orthogonality” requirements are given in the text.

<table>
<thead>
<tr>
<th>Selection</th>
<th>$N_{obs}/N_{bkg}$</th>
<th>$N_{obs}$</th>
<th>$N_{bkg}$</th>
<th>$N_{ggF}$</th>
<th>$N_{VBF}$</th>
<th>$N_{VH}$</th>
<th>$N_{W^{WW}}$</th>
<th>$N_{W^{EW}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{top}}$</th>
<th>$N_{V^{w}}$</th>
<th>$N_{V^{ee}}$</th>
<th>$N_{Drell-Yan}$</th>
<th>$N_{DY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$ category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>0.99 ± 0.00</td>
<td>56759</td>
<td>57180</td>
<td>76</td>
<td>29</td>
<td>24</td>
<td>1330</td>
<td>52020</td>
<td>959</td>
<td>324</td>
<td>324</td>
<td>2550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{tr} &lt; m_{Z} - 25$</td>
<td>1.02 ± 0.01</td>
<td>6777</td>
<td>6650</td>
<td>56</td>
<td>23</td>
<td>15</td>
<td>964</td>
<td>3190</td>
<td>407</td>
<td>233</td>
<td>1850</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VBF orthogonality</td>
<td>1.06 ± 0.02</td>
<td>3826</td>
<td>3620</td>
<td>49</td>
<td>19</td>
<td>12</td>
<td>610</td>
<td>2120</td>
<td>248</td>
<td>152</td>
<td>485</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VH orthogonality</td>
<td>1.05 ± 0.02</td>
<td>3736</td>
<td>3550</td>
<td>44</td>
<td>9.0</td>
<td>12</td>
<td>593</td>
<td>2090</td>
<td>241</td>
<td>148</td>
<td>477</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{T} &lt; 55$</td>
<td>1.04 ± 0.02</td>
<td>3305</td>
<td>3170</td>
<td>40</td>
<td>8.6</td>
<td>7.4</td>
<td>532</td>
<td>1870</td>
<td>212</td>
<td>132</td>
<td>423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi_{T_{T}} &lt; 1.8$</td>
<td>1.09 ± 0.03</td>
<td>1310</td>
<td>1200</td>
<td>35</td>
<td>7.5</td>
<td>5.0</td>
<td>158</td>
<td>572</td>
<td>124</td>
<td>66</td>
<td>282</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{3}{4} m_{T} &lt; m_{T} &lt; m_{H}$</td>
<td>1.06 ± 0.03</td>
<td>1017</td>
<td>955</td>
<td>32</td>
<td>6.9</td>
<td>4.5</td>
<td>140</td>
<td>523</td>
<td>99</td>
<td>60</td>
<td>133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{4} m_{H} &lt; m_{T} &lt; m_{H}$</td>
<td>1.05 ± 0.07</td>
<td>210</td>
<td>200</td>
<td>13.3</td>
<td>2.6</td>
<td>1.9</td>
<td>35</td>
<td>131</td>
<td>16</td>
<td>15</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
independently in the statistical analysis using a fit procedure described in Sec. VII. Inputs to the fit include the event yields and distributions at the final stage of the event selection without the $m_T$ requirement.

The total signal efficiency for $H \rightarrow WW \rightarrow l\nu l\nu$ events produced with $\ell = e$ and $\mu$, including all signal categories and production modes, is 10.2% at 8 TeV for a Higgs boson mass of 125.36 GeV. The corresponding signal efficiency when considering only the VBF production mode is 7.8%.

Figure 11 shows the $m_T$ distributions in the $n_j = 0$, $n_j = 1$, and $n_j \geq 2$ ggF-enriched categories for the 8 TeV data. The distributions for the $n_j \leq 1$ categories are shown in Fig. 12 for the 7 TeV data. The final $O_{BDT}$ output distribution, for the VBF-enriched category, is shown in Fig. 13 for the 7 TeV and 8 TeV data samples.

Figures 14 and 15 show the $p_{T\ell}$ and $m_{\ell\ell}$ distributions at the end of the event selection in the $n_j \leq 1 e\mu$ categories for the 8 TeV data analysis. The distributions are shown for two categories of events based on the flavor of the lepton with the higher $p_T$. This division is important for separating events based on the relative contribution from the backgrounds from misidentified
leptons (W + jets and multijets); see Sec. VI C for details. The dependence of the misidentified lepton and VV background distributions on $p_T^{l_2}$ motivates the separation of the data sample into three bins of $p_T^{l_2}$. The variations in the background composition across the $m_{ll}$ range motivate the division into two bins of $m_{ll}$. Figure 16 shows the corresponding distributions in the $e\mu_{nj \leq 1}$ samples in the 7 TeV data analysis.

The event displays in Fig. 17 show examples of the detector activity for two signal candidates: one in the $n_j = 0$ $e\mu$ category for the 7 TeV data analysis, and one in the VBF-enriched $n_j \geq 2$ $e\mu$ category for the 8 TeV data analysis. Both events have a small value of $\Delta\phi_{ll}$ as is characteristic of the signal. The latter event shows two well-separated jets that are characteristic of VBF production.

V. SIGNAL PROCESSES

The leading Higgs boson production processes are illustrated in Fig. 1. This section details the normalization and simulation of the ggF and VBF production modes. In both cases, the production cross section has been calculated to NNLO in QCD and next-to-leading order in the electroweak couplings. Resummation has been performed to NNLL in QCD for the ggF process. For the decay, the calculation of the branching fraction is computed using the $H \rightarrow WW^*$ and $H \rightarrow ZZ^*$ partial widths from PROPHECY4f [65] and the width of all other decays from HDECAY [66].
The $H \rightarrow WW^*$ branching fraction is 22% with a relative uncertainty of 4.2% for $m_H = 125.36$ GeV \cite{67}. Interference with direct $WW$ production \cite{68} and uncertainties on $VH$ production \cite{69} have a negligible impact on this analysis. Uncertainties on the ggF and VBF production processes are described in the following subsections.

### A. Gluon fusion

The measurement of Higgs boson production via gluon fusion, and the extraction of the associated Higgs boson couplings, relies on detailed theoretical calculations and Monte Carlo simulation. Uncertainties on the perturbative calculations of the total production cross section and of the cross sections exclusive in jet multiplicity are among the leading uncertainties on the expected signal event yield and the extracted couplings. The POWHEG \cite{49} generator matched to PYTHIA8 is used for event simulation and accurately models the exclusive jet multiplicities relevant to this analysis. The simulation is corrected to match higher-order calculations of the Higgs boson $p_T$ distribution.

Production of a Higgs boson via gluon fusion proceeds dominantly through a top-quark loop (the bottom-quark

![ATLAS $H \rightarrow WW^*$](https://example.com/ATLAS.png)

**FIG. 14** (color online). Distributions of the subleading lepton $p_T$ for the 8 TeV data analysis in the $e\mu$ sample used for the statistical analysis described in Sec. VII. The distributions are shown for two categories of events based on the flavor of the subleading lepton $l_2$. The plots are made after requiring all selections up to the $m_T$ requirement, as shown in Tables V and VI. The arrows indicate the bin boundaries; see Fig. 5 for plotting details.

![ATLAS $H \rightarrow WW^*$](https://example.com/ATLAS.png)

**FIG. 15** (color online). Distributions of the dilepton invariant mass $m_{ll}$ for the 8 TeV data analysis in the $e\mu$ sample used for the statistical analysis described in Sec. VII. The distributions are shown for two categories of events based on the flavor of the subleading lepton $l_2$. The plot is made after requiring all selections up to the $m_T$ requirement, as shown in Tables V and VI. The arrows indicate the bin boundaries; see Fig. 5 for plotting details.

---

**OBSERVATION AND MEASUREMENT OF HIGGS BOSON ...**

**PHYSICAL REVIEW D** \textbf{92}, 012006 (2015)
loop contributes 7% to the cross section). Higher-order QCD corrections include radiation from the initial-state gluons and from the quark loop. The total cross section is computed to NNLO [70] using the $m_t \to \infty$ approximation, where an effective pointlike ggH coupling is introduced. Corrections for the finite top-quark mass have been computed to NLO and found to be a few percent [71]; this difference is applied as a correction to the NNLO cross section. The uncertainty on the jet multiplicity distribution is evaluated using the jet-veto-efficiency (JVE) method [76,79] for the ggF categories and the Stewart-Tackmann (ST) method [80] for the VBF category. The JVE method factorizes the total cross section from the acceptances of the jet vetoes in the zero-jet and one-jet channels, treating these components as uncorrelated. Three calculations of the jet veto that defines the zero-jet category, the central value is evaluated at the highest $\sqrt{s}$ (see Sec. IV E). The arrows indicate the bin boundaries; see Fig. 5 for plotting details.

FIG. 16 (color online). Distributions of the subleading lepton $p_T$ and dilepton invariant mass for the 7 TeV data analysis in the $e\mu$ sample. The plots are made after requiring all selections up to $m_T$ (see Sec. IV E). The arrows indicate the bin boundaries; see Fig. 5 for plotting details.

The uncertainty on the jet multiplicity distribution is evaluated using the jet-veto-efficiency (JVE) method [76,79] for the ggF categories and the Stewart-Tackmann (ST) method [80] for the VBF category. The JVE method factorizes the total cross section from the acceptances of the jet vetoes in the zero-jet and one-jet channels, treating these components as uncorrelated. Three calculations of the jet veto that defines the zero-jet category, the central value is evaluated at the highest $\sqrt{s}$ (see Sec. IV E). The arrows indicate the bin boundaries; see Fig. 5 for plotting details.

For the efficiency $\epsilon_0$ of the jet veto that defines the zero-jet channel, the central value is evaluated at the highest available fixed order (NNLO), with NNLL resummation. The uncertainty is taken as the maximum effect of the scale variations on the calculation, or the maximum deviation of the other calculations from this one. The results using the jetveto computation [81] are shown in Fig. 18, along with the reweighted POWHEG + PYTHIA8 prediction.
evaluated with neither hadronization nor the underlying event. The results are consistent to within a few percent for a jet $p_T$ threshold of 25 GeV, and the relative uncertainty at this threshold is 11%.

The efficiency of vetoing an additional jet, given the presence of a single jet, is defined as $\epsilon_1$. The NNLO $n_j \geq 1$ cross section needed for the highest-order calculation of the jet-veto-efficiency method is not available, though the other two calculations of the veto efficiency can be performed using the MCFM generator. The highest-order calculation is in the range spanned by the other two calculations in both the case of $\epsilon_0$ and in the case of $\epsilon_1$ evaluated using a partial calculation of the NNLO $n_j \geq 1$ cross section [82]. The central value of $\epsilon_1$ is thus estimated to be the average of the available calculations, with the uncertainty given by the maximum of the difference with respect to these calculations or the scale-varied estimate of the average. This results in a relative uncertainty of 15% on $\epsilon_1$, as shown in Fig. 18. The figure shows that the reweighted $\text{POWHEG} + \text{PYTHIA8}$ prediction for $\epsilon_1$ agrees

FIG. 17 (color online). Event displays of $H \to WW^{*} \to e\nu\nu\mu$ candidates in the $n_j = 0$ (top) and $n_j \geq 2$ VBF-enriched (bottom) categories. The neutrinos are represented by missing transverse momentum (MET, dotted line) that points away from the $e\mu$ system. The properties of the first event are $p_T^e = 33$ GeV, $p_T^\mu = 24$ GeV, $m_{e\mu} = 48$ GeV, $\Delta\phi_{e\mu} = 1.7$, $p_T^{\text{miss}} = 37$ GeV, and $m_T = 98$ GeV. The properties of the second event are $p_T^e = 51$ GeV, $p_T^\mu = 15$ GeV, $m_{e\mu} = 21$ GeV, $\Delta\phi_{e\mu} = 0.1$, $p_T^{\text{jet}1} = 67$ GeV, $p_T^{\text{jet}2} = 41$ GeV, $m_{jj} = 1.4$ TeV, $\Delta y_{jj} = 6.6$, $p_T^{\text{miss}} = 59$ GeV, and $m_T = 127$ GeV. Both events have a small value of $\Delta\phi_{e\mu}$, which is characteristic of the signal. The second event shows two well-separated jets that are characteristic of VBF production.
with the calculation to within a few percent for a jet $p_T$ threshold of 25 GeV.

A prior ATLAS analysis in this decay channel [5] relied on the ST procedure for all uncertainties associated with jet binning. The JVE estimation reduces uncertainties in the ggF categories by incorporating a resummation calculation (in $\epsilon_0$) and the NLO calculation of $H + 2$ jets (in $\epsilon_1$). The uncertainties for the ST (JVE) procedure are 18% (15%), 43% (27%), and 70% (34%) for the cross sections in the $n_j = 0$, $n_j = 1$, and $n_j \geq 2$ ggF-enriched categories, respectively. These uncertainties are reduced when the categories are combined, and contribute a total of $\approx 5\%$ to the uncertainty on the measured ggF signal strength (see Table XXVI).

Additional uncertainties on the signal acceptance are considered in each signal category. The scale and PDF uncertainties are typically a few percent. A generator uncertainty is taken from a comparison between POWHEG + HERWIG and AMC@NLO + HERWIG [83], which differ in their implementation of the NLO matrix element and the matching of the matrix element to the parton shower. Uncertainties due to the underlying event and parton shower models (UE/PS) are generally small, though in the $n_j = 1$ category they are as large as 14% in the signal regions where $p_T^2 < 20$ GeV. The UE/PS uncertainties are estimated by comparing predictions from POWHEG + HERWIG and POWHEG + PYTHIA8.

The evaluation of the ggF background to the $n_j \geq 2$ VBF category includes an uncertainty on the acceptance of the central-jet veto. The uncertainty is evaluated to be 29% using the ST method, which treats the inclusive $H + 2$-jet and $H + 3$-jet cross sections as uncorrelated. Scale uncertainties are also evaluated in each measurement range of the BDT output, and are 3%–7% in BDT bins 1 and 2, and 48% in BDT bin 3. Other uncertainties on ggF modeling are negligible in this category, except those due to UE/PS, which are significant because the second jet in ggF $H + 2$-jet events is modeled by the parton shower in the POWHEG + PYTHIA8 sample. A summary of the uncertainties on the gluon-fusion and vector-boson-fusion processes is given in Table X. The table shows the uncertainties for same-flavor leptons in the $n_j \leq 1$ categories, since events with different-flavor leptons

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$n_j = 0$</th>
<th>$n_j = 1$</th>
<th>$n_j \geq 2$</th>
<th>$n_j \geq 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gluon fusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cross section</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7.2</td>
</tr>
<tr>
<td>Jet binning or veto</td>
<td>11</td>
<td>25</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>Acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>1.4</td>
<td>1.9</td>
<td>3.6</td>
<td>48</td>
</tr>
<tr>
<td>PDF</td>
<td>3.2</td>
<td>2.8</td>
<td>2.2</td>
<td>-</td>
</tr>
<tr>
<td>Generator</td>
<td>2.5</td>
<td>1.4</td>
<td>4.5</td>
<td>-</td>
</tr>
<tr>
<td>UE/PS</td>
<td>6.4</td>
<td>2.1</td>
<td>1.7</td>
<td>15</td>
</tr>
<tr>
<td>Vector-boson fusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cross section</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.0</td>
</tr>
<tr>
<td>PDF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.0</td>
</tr>
<tr>
<td>Generator</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.2</td>
</tr>
<tr>
<td>UE/PS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14</td>
</tr>
</tbody>
</table>
are further subdivided according to $m_{\ell\ell}$ and $p_T^Z$ (as described in Sec. II).

### B. Vector-boson fusion

The VBF total cross section is obtained using an approximate QCD NNLO computation provided by the vbf@nnlo program [84]. The calculation is based on the structure-function approach [85] that considers the VBF process as two deep-inelastic scattering processes connected to the colorless vector-boson fusion producing the Higgs boson. Leading-order contributions violating this approximation are explicitly included in the computation; the corresponding higher-order terms are negligible [69]. Electroweak corrections are evaluated at NLO with the HAWK program [86]. The calculation has a negligible QCD scale uncertainty and a 2.7% uncertainty due to PDF modeling.

The POWHEG [50] generator is used to simulate the VBF process (see Table III). Uncertainties on the acceptance are evaluated for several sources: the impact of the QCD scale on the jet veto and on the remaining acceptance; PDFs; generator matching of the matrix element to the parton shower; and the underlying event and parton shower. Table X shows the VBF and ggF uncertainties in the most sensitive bin of the BDT output (bin 3). The other bins have the same or similar uncertainties for the VBF process, except for UE/PS, where the uncertainty is 5.2% (<1%) in bin 2 (bin 1).

### VI. BACKGROUND PROCESSES

The background contamination in the various signal regions (SR) comes from several physics processes that were briefly discussed in Sec. II and listed in Table I. They are

(i) $WW$: nonresonant $W$ pair production;
(ii) top quarks (Top): pair production ($tt$) and single-top production ($t$) both followed by the decay $t \rightarrow Wb$;
(iii) misidentified leptons (Misid): $W$ boson production in association with a jet that is misidentified as a lepton ($Wj$) and dijet or multijet production with two misidentifications ($jj$);
(iv) other dibosons (VV): $W\gamma$, $W^*\gamma$, $WZ$ and $ZZ$; and
(v) Drell-Yan (DY): $Z/\gamma^* \rightarrow e\mu$ pairs ($ee/\mu\mu$) and $\tau$ pairs ($\tau\tau$);

the contamination of Higgs decays to non-$WW$ channels is small, but considered as signal. A few background processes, such as $Z\gamma$ and $WW$ produced by double parton interactions, are not listed because their contributions are negligible in the control and signal regions, but they are considered in the analysis for completeness. Their normalizations and acceptances are taken from Monte Carlo simulation.

For each background the event selection includes a targeted set of kinematic requirements (and sample selection) to distinguish the background from the signal. The background estimate is made with a control region (CR) that inverts some or all of these requirements and in many cases enlarges the allowed range for certain kinematic variables to increase the number of observed events in the CR. For example, the relevant selections that suppress the $WW$ background in the $n_j = 0$ SR are $m_{\ell\ell} < 55$ GeV and $\Delta \phi_{\ell\ell} < 1.8$. The WW CR, in turn, is defined by requiring $55 < m_{\ell\ell} < 110$ GeV and $\Delta \phi_{\ell\ell} \leq 2.6$.

The most common use of a CR, like the WW example above, is to determine the normalization factor $\beta$ defined by the ratio of the observed to expected yields of $WW$ candidates in the CR, where the observed yield is obtained by subtracting the non-$WW$ (including the Higgs signal) contributions from the data. The estimate $B_{\text{SR}}^{\text{est}}$ of the expected background in the SR under consideration can be written as

$$B_{\text{SR}}^{\text{est}} = B_{\text{SR}} \cdot \frac{N_{\text{CR}} / N_{\text{CR}}}{B_{\text{SR}} / N_{\text{CR}}} = N_{\text{CR}} \cdot \frac{B_{\text{SR}}}{N_{\text{CR}}}$$

where $N_{\text{CR}}$ and $B_{\text{CR}}$ are the observed yield and the MC estimate in the CR, respectively, and $B_{\text{SR}}$ is the MC estimate in the signal region. The first equality defines the data-to-MC normalization factor in the CR, $\beta$; the second equality defines the extrapolation factor from the CR to the SR, $\alpha$, predicted by the MC. With a sufficient number of events available in the CR, the large theoretical uncertainties associated with estimating the background directly from simulation are replaced by the combination of two significantly smaller uncertainties, the statistical uncertainty on $N_{\text{CR}}$ and the systematic uncertainty on $\alpha$.

When the SR is subdivided for reasons of increased signal sensitivity, as is the case for the $e\mu$ sample for $n_j = 0$, a corresponding $\alpha$ parameter is computed for each of the subdivided regions. The CR (hence the $\beta$ parameter), however, is not subdivided for statistical reasons.

The uncertainties described in this section are inputs to the extraction of the signal strength parameter using the likelihood fit, which is described in Sec. VII. An extension of this method is used when it is possible to determine the extrapolation factor $\alpha$ from data. As described in Secs. VI C and VI E, this can be done for the misidentified lepton backgrounds and in the high-statistics categories for the $Z/\gamma^* \rightarrow ee, \mu\mu$ background. For the former, the distribution of the discriminating variable of interest is also determined from data. For completeness, one should note that the smaller background sources are estimated purely from simulation.

Table XI summarizes, for all the relevant background processes, whether data or MC is used to determine the various aspects of the method. In general, data-derived methods are preferred and MC simulation is used for a few
background processes that do not contribute significantly in the signal region, that have a limited number of events in the control region, or both. MC simulation is used (open circles) or a data sample is used (solid circles) for each of the three aspects of a given method: the normalization (N), the extrapolation (E), and the distribution of the discriminating variable of interest (V). The plots in this section (Figs. 19–27) show contributions that are normalized according to these methods.

This section focuses on the methodology for background predictions and their associated theoretical uncertainties. The experimental uncertainties also contribute to the total uncertainty on these background predictions and are quoted here only for the backgrounds from misidentified leptons, for which the total systematic uncertainties are discussed in Sec. VI C. Furthermore, although the section describes one background estimation technique at a time, the estimates for most background contributions are interrelated and are determined in situ in the statistical part of the analysis (see Sec. VII).

The section is organized as follows. Section VI A describes the WW background in the various categories. This background is the dominant one for the most sensitive \( n_j = 0 \) category. Section VI B describes the background from top-quark production, which is largest in the categories with one or more high-\( p_T \) jets. The data-derived estimate from misidentified leptons is described in Sec. VI C. The remaining backgrounds, VV and \( Z/\gamma^* \), are discussed in Secs. VI D and VI E, respectively. The similarities and modifications for the background estimation for the 7 TeV data analysis are described in Sec. VI F.

Finally, Sec. VI G presents a summary of the background predictions in preparation for the fit procedure described in Sec. VII.

### A. WW dibosons

The nonresonant WW production process, with subsequent decay \( WW \rightarrow \ell \ell \nu \nu \), is characterized by two well-separated charged leptons. By contrast, the charged leptons in the \( H \rightarrow WW^* \rightarrow \ell \ell \nu \nu \) process tend to have a small opening angle (see Fig. 3). The invariant mass of the charged leptons, \( m_{\ell\ell} \), combines this angular information with the kinematic information associated with the relatively low Higgs boson mass \( (m_H < 2m_W) \), providing a powerful discriminant between the processes (see Fig. 7). This variable is therefore used to define WW control regions in the \( n_j \leq 1 \) categories, where the signal is selected with the requirement \( m_{\ell\ell} < 55 \text{ GeV} \). For the \( n_j \geq 2 \) ggF and VBF categories, the WW process is modeled with a merged multiparton SHERPA sample and normalized to the NLO inclusive WW calculation from MCFM [51], since the large top-quark backgrounds make a control-region definition more challenging.

#### 1. \( m_{\ell\ell} \) extrapolation for \( n_j \leq 1 \)

The \( n_j \leq 1 \) analyses use a data-based normalization for the WW background, with control regions defined by a range in \( m_{\ell\ell} \) that does not overlap with the signal regions. The normalization is applied to the combined \( (q\bar{q} \text{ or } gg) \rightarrow WW \) and \( gg \rightarrow WW \) background estimate, and theoretical uncertainties on the extrapolation are evaluated.

To obtain control regions of sufficient purity, several requirements are applied. In order to suppress the \( Z/\gamma^* \) background, the CRs use \( e\mu \) events selected after the \( p_T^{\ell\ell} > 30 \text{ GeV} \) and \( m_{\ell\ell} > 50 \text{ GeV} \) requirements in the \( n_j = 0 \) and \( n_j = 1 \) categories, respectively. The latter requirement additionally suppresses background from multijet production. A requirement of \( p_T^{\ell\ell} > 15 \text{ GeV} \) is applied to suppress the large \( W + \text{jets} \) background below this threshold. Additional \( Z/\gamma^* \rightarrow \tau \tau \) reduction is achieved by requiring \( \Delta \phi_{\ell\ell} < 2.6 \) for \( n_j = 0 \), and \( |m_{\tau\tau} - m_Z| > 25 \text{ GeV} \) for \( n_j = 1 \), where \( m_{\tau\tau} \) is defined in Sec. IV B. The \( m_{\ell\ell} \) range is \( 55 < m_{\ell\ell} < 110 \text{ GeV} \) \( (m_{\ell\ell} > 80 \text{ GeV} \) for \( n_j = 0 \) (1), and is chosen to maximize the signal significance. Increasing the upper bound on \( m_{\ell\ell} \) for \( n_j = 0 \) increases the theoretical uncertainty on the WW background prediction. The \( m_{\tau\tau} \) distributions in the WW control regions are shown in Fig. 19.

The WW estimate \( B_{WW,i}^{\text{est}} \) in each signal region \( i \) is given by Eq. (7). The control region is approximately 70% (45%) pure in the \( n_j = 0 \) (1) category. The contamination in the \( n_j = 1 \) category is dominated by \( t\bar{t} \rightarrow WbWb \) events, where one jet is unidentified and the other is misidentified as a light-quark jet. The single-top contribution is one-third the size of this background for \( n_j = 1 \); for \( n_j = 0 \) this ratio...
is about one-half. All backgrounds are subtracted as part of the fit for $\beta$ described in Sec. VII B 1.

The CR-to-SR extrapolation factor has uncertainties due to the limited accuracy of the MC prediction. Uncertainties due to higher perturbative orders in QCD not included in the MC simulation are estimated by varying the renormalization and factorization scales independently by factors of 1/2 and 2, keeping the ratio of scales in the range 1/2 to 2 [67]. An uncertainty due to higher-order electroweak corrections is determined by reweighting the MC simulation to the NLO electroweak calculation [87] and taking the difference with respect to the nominal sample. PDF uncertainties are evaluated by taking the largest difference between the nominal CT10 [43] PDF set and either the MSTW2008 [88] or the NNPDF2.3 [89] PDF set, and adding in quadrature the uncertainty determined using the CT10 error eigenvectors. Additional uncertainties are evaluated using the same procedures as for ggF production (Sec. VA): uncertainties due to the modeling of the underlying event, hadronization, and parton shower are evaluated by comparing predictions from POWHEG + PYTHIA6 and POWHEG + HERWIG; a generator uncertainty is estimated with a comparison of POWHEG + HERWIG and AMC@NLO + HERWIG. The detailed uncertainties in each signal subregion are given in Table XII. The corresponding uncertainties on the $m_T$ distribution give a relative change of up to 20% between 90 and 170 GeV, depending on the signal region.

The contribution from the $gg \rightarrow WW$ process is 5.8% (6.5%) of the total WW background in the $n_\ell = 0$ (1) category in the signal region and 4.5% (3.7%) in the control region. Its impact on the extrapolation factor is approximately given by the ratio of $gg \rightarrow WW$ to $q\bar{q} \rightarrow WW$ events in the signal region, minus the corresponding ratio in the control region. The leading uncertainty on these ratios is the limited accuracy of the production cross section of the gluon-initiated process, for which a full NLO calculation is not available. The uncertainty evaluated using renormalization and factorization scale variations in the leading-order calculation is 26% (33%) in the $n_\ell = 0$

### TABLE XII. WW theoretical uncertainties (in %) on the extrapolation factor $\alpha$ for $n_\ell \leq 1$. Total is the sum in quadrature of the uncertainties due to the QCD factorization and renormalization scales (Scale), the PDFs, the matching between the hard-scatter matrix element and the UE/PS model (Gen), the missing electroweak corrections (EW), and the parton shower and underlying event (UE/PS). The negative sign indicates anti-correlation with respect to the unsigned uncertainties for SR categories in the same column. Energy-related values are given in GeV.

<table>
<thead>
<tr>
<th>SR category</th>
<th>$n_\ell = 0$</th>
<th>$n_\ell = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$, $10 &lt; m_{\ell\ell} &lt; 30$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_T^{\ell\ell} &gt; 20$</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>$15 &lt; p_T^{\ell\ell} \leq 20$</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td>$10 &lt; p_T^{\ell\ell} \leq 15$</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>$e\mu$, $30 &lt; m_{\ell\ell} &lt; 55$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_T^{\ell\ell} &gt; 20$</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>$15 &lt; p_T^{\ell\ell} \leq 20$</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>$10 &lt; p_T^{\ell\ell} \leq 15$</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>$ee/\mu\mu$, $12 &lt; m_{\ell\ell} &lt; 55$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_T^{\ell\ell} &gt; 10$</td>
<td>0.8</td>
<td>1.1</td>
</tr>
</tbody>
</table>

FIG. 19 (color online). WW control region distributions of transverse mass. The normalizations of all processes are as described in Sec. VI. See Fig. 5 for plotting details.
Boson pairs can be produced by double parton interactions (DPI) in $pp$ collisions. The DPI contribution is very small—$0.4\%$ of WW production in the signal regions—and is estimated using PYTHIA8 MC events normalized to the predicted cross section (rather than the $\beta$ parameter from the WW CR). The cross section is computed using the NNLO $W^\pm$ production cross section and an effective multiparton interaction cross section, $\sigma_{\text{eff}} = 15\text{ mb}$, measured by ATLAS using $Wjj$ production [92]. An uncertainty of $60\%$ is assigned to the value of $\sigma_{\text{eff}}$—and, correspondingly, to the DPI yields—using an estimate of $\sigma_{\text{eff}} \approx 24\text{ mb}$ for WW production [93]. While these estimates rely on theoretical assumptions, an increase of the DPI cross section by a factor of 10 only increases the measured $\mu$ by $1\%$. Background from two $pp \to W$ collisions in the same bunch crossing is negligible.

In the $n_j = 0$ SR, the ratio of signal to WW background is about 1:5, magnifying the impact of background systematic uncertainties. The definition of the CR as a neighboring $m_{\ell\ell}$ window reduces the uncertainty in the extrapolation to low $m_{\ell\ell}$. To validate the assigned uncertainties, the CR normalization is extrapolated to $m_{\ell\ell} > 110\text{ GeV}$ and compared to data. The data are consistent with the prediction at the level of 1.1 standard deviations considering all systematic uncertainties.

The normalization factors determined using predicted and observed event yields are $\rho_{WW}^{\beta} = 1.22 \pm 0.03(\text{stat}) \pm 0.10(\text{syst})$ and $\rho_{WW}^{\beta} = 1.05 \pm 0.05(\text{stat}) \pm 0.24(\text{syst})$, which are consistent with the theoretical prediction at the level of approximately 2 standard deviations. Here the uncertainties on the predicted yields are included though they do not enter into the analysis. Other systematic uncertainties are also suppressed in the full likelihood fit described in Sec. VII B.

2. MC evaluation for $n_j \geq 2$

For the VBF and ggF $n_j \geq 2$ analyses, the WW background is estimated using SHERPA. The MC samples are generated as merged multileg samples, split between the cases where final-state jets result from QCD vertices or from electroweak vertices. The interference between these diagrams is evaluated to be less than a few percent using MADGRAPH; this is included as an uncertainty on the prediction.

For the processes with QCD vertices, uncertainties from higher orders are computed by varying the renormalization and factorization scales in MADGRAPH and are found to be $27\%$ for the VBF category and $19\%$ for the ggF category. Differences between SHERPA and MADGRAPH predictions after selection requirements are $8\%–14\%$ on the $O_{\text{BDT}}$ distribution and $1\%–7\%$ on the $m_T$ distribution, and are taken as uncertainties. The same procedures are used to estimate uncertainties on processes with only electroweak vertices, giving a normalization uncertainty of $10\%$ and an uncertainty on the $O_{\text{BDT}} (m_T)$ distribution of $10\%–16\%$ (5\%–17\%).

The MC prediction is validated using a kinematic selection that provides a reasonably pure sample of $WW\to ll$ events. Events are selected if they pass the preselection requirements on lepton $p_T$ and $m_{\ell\ell}$, have two jets, and $n_b = 0$. An additional requirement of $m_T > 100\text{ GeV}$ is applied in order to enhance the WW contribution. A final discriminant is the minimum of all possible calculations of $m_{T2}$ [94] that use the momenta of a lepton and a neutrino, or the momenta of a lepton, a jet, and a neutrino. The possible momentum values of each neutrino, given $p_T^{\text{miss}}$, are scanned in order to calculate $m_T$; this scan determines $m_{T2}$. A requirement that the minimum $m_{T2}$ be larger than $160\text{ GeV}$ provides a purity of $60\%$ for $WW\to ll$ events (see Fig. 20). The ratio of the observed to the expected number of $WW\to ll$ events in this region is $1.15 \pm 0.19(\text{stat})$.

B. Top quarks

At hadron colliders, top quarks are produced in pairs ($tt$) or in association with a $W$ boson ($Wt$) or quark(s) $q$ (single-$t$). The leptonic decay of the $W$ bosons leads to a final state of two leptons, missing transverse momentum and two $b$-jets (one $b$-jet) in $tt$ ($Wt$) production. The single-$t$ production mode has only one $W$ boson in the final state and the second, misidentified, lepton is produced by a jet. The background from these events is estimated together with the $tt$ and $Wt$ processes in spite of the different lepton production mechanism, but the contribution from these processes to the top-quark background is small. For example, these events are $0.5\%$ of the top-quark
background in the \( n_j = 0 \) category. The top-quark background is estimated using the normalization method, as described in Eq. (7). In the \( n_j = 0 \) category, the SR definition includes a jet veto but the CR has no jet requirements. Because of this, the CR and the SR slightly overlap, but the \( n_j = 0 \) SR is only 3\% of the CR and the expected total signal contamination is less than 1\%, so the effect of the overlap on the results is negligible. In the \( n_j = 1 \) category, the SR definition requires \( n_b = 0 \) but the CR has \( n_b \geq 1 \). In the \( n_j = 2 \) VBF category, the CR is defined requiring one and only one \( b \)-tagged jet. Finally, in the \( n_j = 2 \) ggF category, to reduce the impact of \( b \)-tagging systematic uncertainties, the CR is defined for \( n_b = 0 \), and instead \( m_{\ell\ell} > 80 \) GeV is applied to remove overlap with the SR and minimize the signal contribution.

1. Estimation of jet-veto efficiency for \( n_j = 0 \)

For the \( n_j = 0 \) category, the CR is defined after the preselection missing transverse momentum cut, using only the \( e\mu \) channel, with an additional requirement of \( \Delta R_{\ell\ell} < 2.8 \) to reduce the \( Z/\gamma^* \rightarrow \tau \tau \) background. The CR is inclusive in the number of jets and has a purity of 74\% for top-quark events. The extrapolation parameter \( \alpha \) is the fraction of events with zero reconstructed jets and is derived from the MC simulation.

The value of \( \alpha \) is corrected using data in a sample containing at least one \( b \)-tagged jet. A parameter \( b^{1b} \) is defined as the fraction of events with no additional jets in this region. The ratio \( (\alpha_{\text{data}}^{1b}/\alpha_{\text{MC}}^{1b})^2 \) corrects systematic effects that have a similar impact on the \( b \)-tagged and inclusive regions, such as jet energy scale and resolution. The square is applied to account for the presence of two jets in the Born-level \( \bar{t}t \) production. The prediction can be summarized as

\[
B_{\text{est \ top,0j}} = N_{\text{CR}} \cdot \frac{B_{\text{SR}}}{B_{\text{CR}}} \cdot \left( \frac{\alpha_{\text{data}}^{1b}/\alpha_{\text{MC}}^{1b}}{\gamma^{1b}} \right)^2
\]

where \( N_{\text{CR}} \) is the observed yield in the CR and \( B_{\text{CR}} \) and \( B_{\text{SR}} \) are the estimated yields from MC simulation in the CR and SR, respectively.

Theoretical uncertainties arise from the use of MC-simulated top-quark events in the computation of the ratio \( (\alpha_{\text{data}}^{1b}/\alpha_{\text{MC}}^{1b})^2 \). These uncertainties include variations of the renormalization and factorization scales, choice of PDFs, and the parton shower model. The procedure is sensitive to the relative rates of \( Wt \) and \( \bar{t}t \) production, so an uncertainty is included on this cross section ratio and on the interference between these processes. An additional theoretical uncertainty is evaluated on the efficiency \( \epsilon_{\text{rest}} \) of the additional selection after the \( n_j = 0 \) preselection, which is estimated purely from MC simulation. Experimental uncertainties are also evaluated on the simulation-derived components of the background estimate, with the main contributions coming from jet energy scale and resolution. The uncertainties on \( \alpha_{\text{MC}}^{1b}/(\alpha_{\text{MC}}^{1b})^2 \) and on \( \epsilon_{\text{rest}} \) are summarized in Table XIII.

The resulting normalization factor is \( \rho_{\text{top}} = 1.08 \pm 0.02 \) (stat), including the correction factor \( (\alpha_{\text{data}}^{1b}/\alpha_{\text{MC}}^{1b})^2 = 1.006 \). The total uncertainty on the background yield in the \( n_j = 0 \) signal region is 8\%.

2. Extrapolation from \( n_j = 1 \) for \( n_j = 1 \)

In the \( n_j = 1 \) SR, top-quark production is the second leading background, after nonresonant \( WW \) production. Summing over all signal regions with no \( m_{\ell\ell} \) requirement applied, it is 36\% of the total expected background and the ratio of signal to top-quark background is approximately 0.2. It also significantly contaminates the \( n_j = 1 \) WW CR with a yield as large as that of nonresonant WW in this CR.

Two parameters are defined for the extrapolation from the top CR, one to the SR \( (\alpha_{\text{SR}}) \) and one to the WW CR \( (\alpha_{\text{WW}}) \).

The top CR is defined after the preselection in the \( e\mu \) channel and requires the presence of exactly one jet, which must be \( b \)-tagged. There can be no additional \( b \)-tagged jet with \( 20 < p_T < 25 \) GeV, following the SR requirement. The requirement \( m_{\ell\ell} > 50 \) GeV is also applied to reject \( jj \) background. As in the WW case, only the \( e\mu \) events are used in order to suppress the \( Z/\gamma^* \) contamination. The \( m_{\ell\ell} \) distribution in this control region is shown in Fig. 21(a).

The CR requires at least one \( b \)-jet, but the SR requires zero. In the case of a simple extrapolation using the ratio of

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>( \alpha_{\text{MC}}^{1b}/(\alpha_{\text{MC}}^{1b})^2 )</th>
<th>( \epsilon_{\text{rest}} )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) ( n_j = 0 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.4</td>
<td>1.2</td>
<td>4.6</td>
</tr>
<tr>
<td>Non-top-quark subtraction</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Theoretical</td>
<td>3.9</td>
<td>4.5</td>
<td>4.9</td>
</tr>
<tr>
<td>Statistical</td>
<td>2.2</td>
<td>0.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Total</td>
<td>6.8</td>
<td>4.7</td>
<td>7.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regions</th>
<th>Scale</th>
<th>PDF</th>
<th>Gen</th>
<th>UE/PS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) ( n_j = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( e\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 10 &lt; m_{\ell\ell} &lt; 55 )</td>
<td>-1.1</td>
<td>-0.12</td>
<td>-2.4</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td>( e\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 12 &lt; m_{\ell\ell} &lt; 55 )</td>
<td>-1.0</td>
<td>-0.12</td>
<td>-2.0</td>
<td>3.0</td>
<td>3.7</td>
</tr>
<tr>
<td>WW control region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( e\mu ) (( m_{\ell\ell} &gt; 80 ))</td>
<td>0.6</td>
<td>0.08</td>
<td>2.0</td>
<td>1.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>

As shown in Table XII for column headings.
The efficiency measurement \( \epsilon_{2j}^{\text{data}} \) is extrapolated from the \( n_j = 2 \) sample to the \( n_j = 1 \) samples using \( \gamma_{1j} = \epsilon_{1j}/\epsilon_{2j} \), which is evaluated using MC simulation. Jets in the \( n_j = 2 \) and jets in the \( n_j = 1 \) samples have similar kinematic features; one example, the jet \( p_T \), is illustrated in Fig. 21(b). In this figure, the \( n_j = 2 \) distribution contains the \( p_T \) of one of the two jets, chosen at random, provided that the other jet is tagged, so that the distribution contains the same set of jets as is used in the extrapolation to \( n_j = 1 \). Residual disagreements between the distributions are reflected in the deviation of \( \gamma_{1j} \) from unity, which is small. The value of \( \gamma_{1j} = 1.079 \pm 0.002(\text{stat}) \) with an experimental uncertainty of 1.4% and a theoretical uncertainty of 0.8%. The experimental uncertainty is almost entirely due to uncertainties on the \( b \)-tagging efficiency. The theoretical uncertainty is due to the PDF model, renormalization and factorization scales, matching of the matrix element to the parton shower, top-quark cross sections, and interference between top-quark single and pair production.

The estimated \( b \)-tagging efficiency in the \( n_j = 1 \) data is \( \epsilon_{1j}^{\text{est}} = \gamma_{1j} \cdot \epsilon_{2j}^{\text{data}} \) and the top-quark background estimate in the SR is then

\[
B_{\text{top,1j}}^{\text{est}} = N_{\text{CR}} \cdot \left( \frac{1 - \epsilon_{1j}^{\text{est}}}{\epsilon_{1j}^{\text{est}}} \right). \tag{9}
\]

The theoretical systematic uncertainties are summarized in Table XIII. The normalization factor for this background is \( \beta_{\text{top}}^{1j} = 1.06 \pm 0.03(\text{stat}) \), and the total uncertainty on the estimated background in the \( n_j = 1 \) signal region is 5%.

3. Extrapolation from \( n_b = 1 \) for VBF-enriched \( n_j \geq 2 \)

The \( n_j \geq 2 \) categories have a large contribution from top-quark background events even after selection requirements, such as the \( b \)-jet veto, that are applied to reduce them, because of the two \( b \) quarks in \( t\bar{t} \) events. The majority of the residual top-quark events have a light-quark jet from initial-state radiation and a \( b \)-quark jet that is not identified by the \( b \)-tagging algorithm. The CR requires exactly one \( b \)-tagged jet to mimic this topology, so that at first order the CR-to-SR extrapolation factor \( (a) \) is the ratio of \( b \)-jet efficiency to \( b \)-jet inefficiency. The CR includes events from \( ee \) and \( Z/\gamma^* \) final states because the \( Z/\gamma^* \) contamination is reduced by the jet selection.

The \( O_{\text{BDT}} \) discriminant is based on variables, such as \( m_{jj} \), that depend on the jet kinematics, so the acceptance for top-quark events in each \( O_{\text{BDT}} \) bin is strongly dependent on the Monte Carlo generator and modeling. Motivated by the large variation of top-quark event kinematics as a function of the \( O_{\text{BDT}} \) bin, the top-quark background is normalized independently in each bin, which reduces the modeling.
uncertainties. Figure 22 shows the $m_{jj}$ and $O_{\text{BDT}}$ distributions in the top CR used for the VBF category. The two bins with the highest $O_{\text{BDT}}$ score are merged to improve the statistical uncertainty on the estimated background. The uncertainties on the extrapolation from the single bin in the CR to the two bins in the SR are separately evaluated.

Table XIV shows the normalization factors $\beta_i$ and their uncertainties for each $O_{\text{BDT}}$ bin, as well as the theoretical uncertainties on the extrapolation factors $\alpha_j$ to the corresponding SR bins.

The uncertainties on $\alpha$ were evaluated with the same procedure used for the WW background (see Sec. VI A 1). The only significant source is a modeling uncertainty evaluated by taking the maximum spread of predictions from POWHEG+HERWIG, ALPGEN+HERWIG, and MC@NLO+HERWIG [95]. The generators are distinguished by the merging of LO matrix-element evaluations of up to three jets produced in association with $t\bar{t}$ (ALPGEN+HERWIG) or by differences in procedures for matching a NLO matrix-element calculation to the parton shower (MC@NLO+HERWIG and POWHEG+HERWIG). The systematic uncertainty is dominated by the ALPGEN+HERWIG–MC@NLO+HERWIG difference, and for this reason the theoretical uncertainties shown in Table XIV are fully correlated between bins.

### 4. Extrapolation in $m_{T\ell}$ for ggF-enriched $n_j \geq 2$

In the more inclusive phase space of the ggF-enriched $n_j \geq 2$ category, the $t\bar{t}$ background remains dominant after the $n_k = 0$ requirement, as is the case for the VBF-enriched category. The CR is defined with $m_{T\ell} > 80$ GeV to distinguish it from the signal region (see Fig. 10) and reduce signal contamination. The CR is approximately 70% pure in top-quark events, and a normalization factor of $\beta = 1.05 \pm 0.03 (\text{stat})$ is obtained. The uncertainties on the extrapolation factor $\alpha$ to the SR are 3.2% from the comparison of MC@NLO+HERWIG, ALPGEN+HERWIG, and POWHEG+PYTHIA; 1.2% for the parton shower and underlying-event uncertainties from the comparison of POWHEG+PYTHIA6 and POWHEG+HERWIG; 1% from the missing higher-order contribution, evaluated by varying the renormalization and factorization scales; 0.3% from the PDF envelope evaluated as described in Sec. VI A 1; and 0.7% from the experimental uncertainties. The effect of the same set of variations on the predicted $m_{T\ell}$ distribution in the signal region was also checked. The variations from the nominal distribution are small, at most 4% in the tails, but they are included as a shape systematic in the fit procedure.

### C. Misidentified leptons

Collisions producing $W$ bosons in association with one or more jets—referred to here as $W + \text{jets}$—may enter the signal sample when a jet is misidentified as a prompt lepton. In this background, there is a prompt lepton and a transverse momentum imbalance from the leptonic decay of the $W$ boson. Background can also arise from multijet production when two jets are misidentified...
as prompt leptons and a transverse momentum imbalance is reconstructed.

I. $W + \text{jets}$

The $W + \text{jets}$ background contribution is estimated using a control sample of events where one of the two lepton candidates satisfies the identification and isolation criteria for the signal sample, and the other lepton fails to meet these criteria but satisfies less restrictive criteria (these lepton candidates are denoted “anti-identified”). Events in this sample are otherwise required to satisfy all of the signal selection requirements. The dominant component of this sample (85% to 90%) is due to $W + \text{jets}$ events in which a jet produces an object reconstructed as a lepton. This object may be either a nonprompt lepton from the decay of a hadron containing a heavy quark, or else a particle (or particles) from a jet reconstructed as a lepton candidate.

The $W + \text{jets}$ contamination in the signal region is obtained by scaling the number of events in the data control sample by an extrapolation factor. This extrapolation factor is measured in a data sample of jets produced in association with $Z$ bosons reconstructed in either the $e^+e^-$ or $\mu^+\mu^-$ final state (referred to as the $Z + \text{jets}$ control sample below). The factor is the ratio of the number of identified lepton candidates satisfying all lepton selection criteria to the number of anti-identified leptons measured in bins of anti-identified lepton $p_T$ and $\eta$. Anti-identified leptons must explicitly not satisfy the signal selection criteria (so that leptons counted in the numerator of this ratio exclude the anti-identified leptons counted in the denominator of this ratio) and the signal requirements for isolation and track impact parameters are either relaxed or removed. In addition, for anti-identified electrons the identification criteria specifically targeting conversions are removed and the anti-identified electron is explicitly required to fail the “medium” electron identification requirement specified in Ref. [23].

Figure 23 shows the $p_T$ distributions of identified muons [Fig. 23(a)], identified electrons [Fig. 23(b)], anti-identified muons [Fig. 23(c)], and anti-identified electrons [Fig. 23(d)] in the $Z + \text{jets}$ control sample. The extrapolation factor in a given $p_T$ bin is the number of identified leptons divided by the number of anti-identified leptons in that particular bin. Each number is corrected for the presence of processes not due to $Z + \text{jets}$. The $Z + \text{jets}$ sample is contaminated by other production processes that produce additional prompt leptons (e.g., $WZ \rightarrow \ell\nu\ell\ell$) or nonprompt leptons not originating from jets (e.g., $Z/\gamma'$ and $Z\gamma$) that create a bias in the extrapolation factor. Kinematic criteria suppress about 80% of the contribution from these other processes in the $Z + \text{jets}$ sample. The remaining total contribution of these other processes after applying these kinematic criteria is shown in the histograms in Fig. 23. The uncertainty shown in these histograms is the 10% systematic uncertainty assigned to the contribution from these other processes, mainly due to cross section uncertainties. This remaining contribution from other processes is estimated using Monte Carlo simulation and removed from the event yields before calculating the extrapolation factor.

The composition of the associated jets—namely the fractions of jets due to the production of heavy-flavor quarks, light-flavor quarks, and gluons—in the $Z + \text{jets}$ sample and the $W + \text{jets}$ sample may be different. Any difference would lead to a systematic error in the estimate

FIG. 23. Misidentified lepton sample distributions of $p_T$ in the $Z + \text{jets}$ control sample: (a) identified muon, (b) identified electron, (c) anti-identified muon, and (d) anti-identified electron. The symbols represent the data (Obs); the histograms are the background MC estimates (Bkg) of the sum of electroweak processes other than the associated production of a $Z$ boson and jets.
of the $W + \text{jets}$ background due to applying the extrapolation factor determined with the $Z + \text{jets}$ sample to the $W + \text{jets}$ control sample, so Monte Carlo simulation is used to determine a correction factor that is applied to the extrapolation factors determined with the $Z + \text{jets}$ data sample. A comparison of the extrapolation factors determined with the $Z + \text{jets}$ sample and the $W + \text{jets}$ sample is made for three Monte Carlo simulations: ALPGEN+PYTHIA6, ALPGEN+HERWIG, and POWHEG+PYTHIA8. For each combination of matrix-element and parton-shower simulations, a ratio of the extrapolation factors for $W + \text{jets}$ versus $Z + \text{jets}$ is calculated. These three ratios are used to determine a correction factor and an uncertainty that is applied to the extrapolation factors determined with the $Z + \text{jets}$ data: this correction factor is $0.99 \pm 0.20$ for anti-identified electrons and $1.00 \pm 0.22$ for anti-identified muons.

The total uncertainties on the corrected extrapolation factors are summarized in Table XV. In addition to the systematic uncertainty on the correction factor due to the sample composition, the other important uncertainties on the $Z + \text{jets}$ extrapolation factor are due to the limited number of jets that meet the lepton selection criteria in the $Z + \text{jets}$ control sample and the uncertainties on the contributions from other physics processes in the identified and anti-identified lepton samples. The total systematic uncertainty on the corrected extrapolation factors varies as a function of the $p_T$ of the anti-identified lepton; this variation is from 29% to 61% for anti-identified electrons and 25% to 46% for anti-identified muons. The systematic uncertainty on the corrected extrapolation factor dominates the systematic uncertainty on the $W + \text{jets}$ background.

<table>
<thead>
<tr>
<th>SR $p_T$ range (GeV)</th>
<th>Total Corr. factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC</td>
</tr>
<tr>
<td><strong>Electrons</strong></td>
<td></td>
</tr>
<tr>
<td>10–15</td>
<td>29</td>
</tr>
<tr>
<td>15–20</td>
<td>44</td>
</tr>
<tr>
<td>20–25</td>
<td>61</td>
</tr>
<tr>
<td>≥25</td>
<td>43</td>
</tr>
<tr>
<td><strong>Muons</strong></td>
<td></td>
</tr>
<tr>
<td>10–15</td>
<td>25</td>
</tr>
<tr>
<td>15–20</td>
<td>37</td>
</tr>
<tr>
<td>20–25</td>
<td>37</td>
</tr>
<tr>
<td>≥25</td>
<td>46</td>
</tr>
</tbody>
</table>

The uncertainties on the signal strength $\mu$ are classified into experimental, theoretical, and other components, as described in Sec. IX and Table XXVI. The uncertainty on $\mu$ due to the correction factor applied to the extrapolation factor is classified as theoretical because the uncertainty on the correction factor is derived from a comparison of predictions from different combinations of Monte Carlo generators and parton shower algorithms. The uncertainty on $\mu$ due to the other uncertainties on the extrapolation factor ($Z + \text{jet}$ control sample statistics and the subtraction of other processes from this control sample) is classified as experimental.

Figure 24 shows the extrapolation factor measured in the $Z + \text{jets}$ data compared to the predicted extrapolation factor determined using Monte Carlo simulated samples...
(ALPGEN + PYTHIA6) of $Z +$ jets and $W +$ jets for anti-identified muons [Fig. 24(a)] and anti-identified electrons [Fig. 24(b)]. The values of the extrapolation factors are related to the specific criteria used to select the anti-identified leptons and, as a result, the extrapolation factor for anti-identified muons is about one order of magnitude larger than the extrapolation factor for anti-identified electrons. This larger extrapolation factor does not indicate a larger probability for a jet to be misidentified as a muon compared to an electron. In fact, misidentified electrons contribute a larger portion of the $W +$ jets background in the signal region.

The $W +$ jets background in the signal region is determined using a control sample in which the lepton and the anti-identified lepton are required to have opposite charge. A prediction of the $W +$ jets background is also used for a data control sample consisting of events that satisfy all of the Higgs boson signal requirements except that the two lepton candidates are required to have the same charge. This same-charge control region is described in Sec. VI D.

The $W +$ jets process is not expected to produce equal numbers of same-charge and opposite-charge candidates. In particular, associated production processes such as $Wc$, where the second lepton comes from the semileptonic decay of a charmed hadron, produce predominantly opposite-charge candidates. Therefore, a separate extrapolation factor is applied to the same-charge $W +$ jets control sample.

The procedure used to determine the same-charge extrapolation factor from the $Z +$ jets data is the same as the one used for the signal region. Because of the difference in jet composition of the same-charge $W +$ jets control sample, a different correction factor is derived from MC simulation to correct the extrapolation factor determined with the $Z +$ jets data sample for application to the same-charge $W +$ jets sample. Figure 24 compares the extrapolation factors in same-charge $W +$ jets with the ones in $Z +$ jets. The correction factor is $1.25 \pm 0.31$ for anti-identified electrons and $1.40 \pm 0.49$ for anti-identified muons; as with the opposite-charge correction factors, these factors and their systematic uncertainty are determined by comparing the factors determined with the three different samples of MC simulations mentioned previously in the text (ALPGEN+PYTHIA6, ALPGEN+HERWIG, and POWHEG+PYTHIA8). The total uncertainties on the corrected extrapolation factors used to estimate the $W +$ jets background in the same-charge control region are shown in Table XV. The correlation between the systematic uncertainties on the opposite-charge and same-charge correction factors reflects the composition of the jets producing objects misidentified as leptons. These jets have a component that is charge symmetric with respect to the charge of the $W$ boson as well as a component unique to opposite-charge $W +$ jets processes. Based on the relative rates of same- and opposite-charge $W +$ jets events, 60% of the opposite-charge correction factor uncertainty is correlated with 100% of the corresponding same-charge uncertainty.

2. Multijets

The background in the signal region due to multijets is determined using a control sample that has two anti-identified lepton candidates, but otherwise satisfies all of the signal region selection requirements. A separate extrapolation factor—using a multijet sample—is measured for the multijet background and applied twice to this control sample. The sample used to determine the extrapolation factor is expected to have a similar sample composition (in terms of heavy-flavor jets, light-quark jets, and gluon jets) to the control sample. Since the presence of a misidentified lepton in a multijet sample influences the sample composition—for example by increasing the fraction of heavy-flavor processes in the multijet sample—corrections to the extrapolation factor are made that take into account this correlation. The event-by-event corrections vary between 1.0 and 4.5 depending on the lepton flavor and $p_T$ of both misidentified leptons in the event; the electron extrapolation factor corrections are larger than the muon extrapolation factor corrections.

3. Summary

Table XVI lists the estimated event counts for the multijet and $W +$ jets backgrounds in the $e\mu$ channel for the various jet multiplicities. The values are given before the $m_T$ fit for the ggF-enriched categories and after the VBF-selection for the VBF-enriched categories. The uncertainties are the combination of the statistical and systematic uncertainties and are predominantly systematic. The dominant systematic uncertainty is from the extrapolation factors. In the case of the $W +$ jets background, these uncertainties are summarized in Table XV; in the case of the multijet background, the largest contribution is the uncertainty introduced by the correlations between extrapolation factors in an event with two misidentified leptons.

<table>
<thead>
<tr>
<th>Category</th>
<th>$W +$ jets yield $N_{Wj}$</th>
<th>Multijets yield $N_{jj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC</td>
<td>SC</td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>278 ± 71</td>
<td>174 ± 54</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>88 ± 22</td>
<td>62 ± 18</td>
</tr>
<tr>
<td>$n_j \geq 2$ ggF &amp; 50 ± 22 &amp; - &amp; 49 ± 22 &amp; -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j \geq 2$ VBF &amp; 3.7 ± 1.2 &amp; - &amp; 2.1 ± 0.8 &amp; -</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE XVI. $W +$ jets and multijets estimated yields in the $e\mu$ category. For $n_j = 0$ and $n_j = 1$, yields for both the opposite-charge and same-charge leptons are given. The yields are given before the $m_T$ fit for the ggF-enriched categories and after the VBF selection for the VBF-enriched categories. The uncertainties are from a combination of statistical and systematic sources.
For the $n_j = 0$ and $n_j = 1$ categories, the expected backgrounds are provided for both the opposite-charge signal region and the same-charge control region (described in Sec. VI D), and the multijet background is expected to be less than 10% of the $W +$ jets background in these two categories. For higher jet multiplicities, the multijet background is expected to be comparable to the $W +$ jets background because there is no selection criterion applied to $m_j^T$. In this case, however, the multijet background has a very different $m_j^T$ distribution than the Higgs boson signal, so it is not necessary to suppress this background to the same extent as in the lower jet multiplicity categories.

D. Other dibosons

There are backgrounds that originate from the production of two vector bosons other than $WW$. These include $W\gamma$, $W\gamma^*$, $WZ$, and $ZZ$ production and are referred to here as $VV$. The $VV$ processes add up to about 10% of the total estimated background in the $n_j \leq 1$ channels and are of the same magnitude as the signal. The dominant sources of these backgrounds are the production of $W\gamma$ and $W\gamma^*/WZ$, where this latter background is a combination of the associated production of a $W$ boson with a nonresonant $Z/\gamma^*$ or an on-shell $Z$ boson.

The normalization of the $VV$ background processes in the $e\mu$ channel is determined from the data using a same-charge control region, which is described below. The distribution of these various contributing processes in the different signal bins is determined using MC simulation. In the $ee/\mu\mu$ channels, both the normalization and the distributions of the $VV$ processes are estimated with MC simulation. The details of these simulations are provided in Sec. III C.

Several specialized data sample selections are used to validate the simulation of the rate and the shape of the various contributing processes and the simulation of the efficiency for rejecting electrons from photon conversions. The $W\gamma$ background enters the signal region when the $W$ boson decays leptonically and the photon converts into an $e^+e^−$ pair in the detector material. If the pair is very asymmetric in $p_T$, then it is possible that only the electron or positron satisfies the electron selection criteria, resulting in a Higgs boson signal candidate. This background has a prompt electron or muon and missing transverse momentum from the $W$ boson decay and a nonprompt electron or positron. The prompt lepton and the conversion product are equally likely to have opposite electric charge (required in the signal selection) and the same electric charge, since the identification is not charge dependent.

A sample of nonprompt electrons from photon conversions can be selected by reversing two of the electron signal selection requirements: the electron track should be part of a reconstructed photon conversion vertex candidate and the track should have no associated hit on the innermost layer of the pixel detector. Using these two reversed criteria, a sample of $e\mu$ events that otherwise satisfy all of the kinematic requirements imposed on Higgs boson signal candidates is selected; in the $n_j = 0$ category ($n_j = 1$ category), 83% (87%) of this sample originates from $W\gamma$ production. This sample is restricted to events selected online with a muon trigger to avoid biases in the electron selection introduced by the online electron trigger requirements. Figures 25(a) and 25(b) show the $m_j^T$ distribution and the $p_T$ distribution of the electron of the $n_j = 0$ category of this $W\gamma$ validation sample compared to expectations from the MC simulation. Verifying that the simulation correctly models the efficiency of detecting photon conversions is important to ensure that the $W\gamma$ background normalization and distributions are accurately modeled. To evaluate the modeling of photon conversions, a $Z \rightarrow \mu\mu\gamma$ validation sample consisting of either $Z\gamma$ or $Z$ boson production with final-state radiation is selected. The $Z$ boson is reconstructed in the $\mu^+\mu^-\gamma$ decay channel, and an electron (or positron) satisfying all the electron selection criteria except the two reversed criteria specified above is selected. The $\mu^+\mu^-e^\pm$ invariant mass is required to be between 15 GeV of $m_Z$ to reduce contributions from the associated production of a $Z$ boson and hadronic jets. The resulting data sample is more than 99% pure in the $Z \rightarrow \mu\mu\gamma$ process. A comparison between this data sample and a $Z \rightarrow \mu\mu\gamma$ MC simulation indicates some potential mismodeling of the rejection of nonprompt electrons in the simulation. Hence a $p_T$-dependent systematic uncertainty ranging from 25% for $10 < p_T < 15$ GeV to 5% for $p_T > 20$ GeV is assigned to the efficiency for nonprompt electrons from photon conversions to satisfy the rejection criteria.

The $W\gamma^*$ background originates from the associated production of a $W$ boson that decays leptonically and a virtual photon $\gamma^*$ that produces an $e^+e^−$ or $\mu^+\mu^−$ pair in which only one lepton of the pair satisfies the lepton selection criteria. This background is most relevant in the $n_j = 0$ signal category, where it contributes a few percent of the total background and is equivalent to about 25% of the expected Higgs boson signal.

The $W\gamma^*$ background is studied with a specific selection aimed at isolating a sample of $W\gamma^* \rightarrow e\nu\mu\mu$ candidates. Events with an electron and a pair of opposite-charge muons are selected with $m_{\mu\nu} < 7$ GeV, $p_T^{\mathrm{miss}} > 20$ GeV and both muons must satisfy $\Delta\phi(e, \mu) < 2.8$. Muon pairs consistent with originating from the decay of a $J/\psi$ meson are rejected. The electron and the highest $p_T$ muon are required to satisfy the signal region lepton selection criteria and $p_T$ thresholds; however, the subleading-muon $p_T$ threshold is reduced to 3 GeV. The isolation criteria for the higher-$p_T$ muon are modified to take into account the presence of the lower-$p_T$ muon. The SHERPA $W\gamma^*$ simulation sample with $m_T < 7$ GeV is compared to the data selected with the above criteria; the
distributions of the $m_T$ calculated using the electron and the higher-$p_T$ muon and the invariant mass of the two muons $m_{\mu\mu}$ are shown in Figs. 25(c) and 25(d).

The WZ and ZZ backgrounds are modeled with MC simulation. No special samples are selected to validate the simulation of these processes. The ZZ background arises primarily when one Z boson decays to $e^+e^-$ and the other to $\mu^+\mu^-$ and an electron and a muon are not detected. This background is very small, amounting to less than 3% of the VV background. Background can also arise from $Z\gamma^*$ and $Z\gamma$ production if the Z boson decays to $\ell^+\ell^-$ and one of the leptons is not identified and the photon results in a second lepton. These backgrounds are also very small, and the $Z\gamma^*$ background is neglected.

The VV backgrounds arising from $W\gamma$, $W\gamma^*$, and $WZ$ are equally likely to result in a second lepton that has the same charge or opposite charge compared to the lepton from the $W$ boson decay. For this reason, a selection of $e\mu$ events that is identical to the Higgs boson candidate selection except that it requires the two leptons to have the same charge is used to define a same-charge control region. The same-charge control region is dominated by VV processes. The other processes that contribute significantly to the same-charge sample are the $W + jets$ process and—to a much lesser extent—the multijet process. The same-charge data sample can be used to normalize the VV processes once the contribution from the $W + jets$ process is taken into account, using the method described in Sec. VI C.

Figure 26 shows the distributions of the transverse mass [26(a) and 26(c)] and the subleading lepton $p_T$ [26(b) and 26(d)] for the same-charge data compared with the MC simulations after normalizing the sum of these MC predictions to the same-charge data. A single normalization factor is applied simultaneously to all four MC simulations of the VV backgrounds (shown separately in the figures). These normalization factors are $\hat{\beta}_{ij} = 0.92 \pm 0.07$ (stat) and $\hat{\beta}_{ij} = 0.96 \pm 0.12$ (stat) for the $e\mu$ channels in the $n_j \leq 1$ categories. The VV processes comprise about 60% of the total in both the zero-jet and one-jet same-charge data samples, with 30% coming from the $W + jets$ process.

Theoretical uncertainties on the VV backgrounds are dominated by the scale uncertainty on the prediction for each jet bin. For the $W\gamma$ process, a relative uncertainty of 6% on the total cross section is correlated across jet categories, and the uncorrelated jet-bin uncertainties are 9%, 53%, and 100% in the $n_j = 0$, $n_j = 1$, and $n_j \geq 2$ categories, respectively. For the $W\gamma^*$ process, the corresponding uncertainties are 7% (total cross section), 7% ($n_j = 0$), 30% ($n_j = 1$), and 26% ($n_j \geq 2$). No uncertainty is applied for the extrapolation of these backgrounds from the same-charge control region to the opposite-charge signal region, since it was verified in the simulation that
these processes contribute equal numbers of opposite-charge and same-charge events.

### E. Drell-Yan

The DY processes produce two oppositely charged leptons and some events are reconstructed with significant missing transverse momentum. This is mostly due to neutrinos produced in the Z boson decay in the case of the $Z/\gamma^* \rightarrow \tau\tau$ background to the $ee/\mu\mu$ channels. In contrast, in the case of the $Z/\gamma^* \rightarrow ee, \mu\mu$ background to the $ee/\mu\mu$ channels, it is mostly due to detector resolution that is degraded at high pile-up and to neutrinos produced in $b$-hadron or $c$-hadron decays (from jets produced in association with the Z boson). Preselection requirements, such as the one on $p_T^{miss}$, reduce the bulk of this background, as shown in Fig. 5, but the residual background is significant in all categories, especially in the $ee/\mu\mu$ samples. The estimation of the $Z/\gamma^* \rightarrow \tau\tau$ background for the $ee$ samples is done using a control region, which is defined in a very similar way across all $n_j$ categories, as described below. Since a significant contribution to the $Z/\gamma^* \rightarrow ee, \mu\mu$ background to the $ee/\mu\mu$ categories arises from mismeasurements of the missing transverse momentum, more complex data-derived approaches are used to estimate this background, as described below.

Mismodeling of $p_T^{\tau/\gamma}$, reconstructed as $p_T^{\text{eff}}$, was observed in the $Z/\gamma^*$-enriched region in the $n_j = 0$ category. The ALPGEN + HERWIG MC generator does not adequately model the parton shower of soft jets that balance $p_T^{\tau/\gamma}$ when there are no selected jets in the event. A correction, based on the weights derived from a data-to-MC comparison in the $Z$ peak, is therefore applied to MC events in the $n_j = 0$ category, for all leptonic final states from Drell-Yan production.

### 1. $Z/\gamma^* \rightarrow \tau\tau$

The $Z/\gamma^* \rightarrow \tau\tau$ background prediction is normalized to the data using control regions. The contribution from this background process is negligible in the $ee/\mu\mu$ channel, and in order to remove the potentially large $Z/\gamma^* \rightarrow ee, \mu\mu$ contamination, the CR is defined using the $ee$ samples in all categories except the $n_j \geq 2$ VBF-enriched one.

The control region in the $n_j = 0$ category is defined by the requirements $m_{ee} < 80$ GeV and $\Delta\phi_{ee} > 2.8$, which select a 91%-pure region and result in a normalization factor $\beta_{0j} = 1.00 \pm 0.02$ (stat). In the $n_j = 1$ category, the invariant mass of the $\tau\tau$ system, calculated with the collinear mass approximation, and defined in Sec. IV B, can be used since the dilepton system is boosted. An 80%-pure region is selected with $m_{\ell\ell} < 80$ GeV and $m_{\tau\tau} > (m_Z - 25$ GeV). The latter requirement ensures that there is no overlap with the signal region selection. The resulting normalization factor is $\beta_{1j} = 1.05 \pm 0.04$ (stat). The $n_j \geq 2$ ggF-enriched category uses a CR selection of $m_{ee} < 70$ GeV and $\Delta\phi_{ee} > 2.8$ providing 74% purity and a normalization factor $\beta_{2j} = 1.00 \pm 0.09$ (stat). Figure 27 shows the $m_\ell$ distributions in the control regions in the $n_j = 0$ and $n_j = 1$ categories. High purity and good data/MC agreement is observed.

In order to increase the available statistics in the $Z/\gamma^* \rightarrow \tau\tau$ control region in the $n_j \geq 2$ VBF-enriched category, $ee/\mu\mu$ events are also considered. The contribution from $Z/\gamma^* \rightarrow ee, \mu\mu$ decays is still negligible. The control region
is defined by the invariant mass requirements: $m_{\ell\ell} < 80$ GeV (75 GeV in $e\mu$) and $|m_{\tau\tau} - m_{Z}| < 25$ GeV. The resulting normalization factor is derived after summing all three bins in $O_{BDT}$ and yields $\beta = 0.9 \pm 0.3\text{(stat)}$.

Three sources of uncertainty are considered on the extrapolation of the $Z/\gamma^* \rightarrow \tau\tau$ background from the control region: QCD scale variations, PDFs, and generator modeling. The latter are evaluated based on a comparison of ALPGEN + HERWIG and ALPGEN + PYTHIA generators. An additional uncertainty on the $p_T^{\gamma^*}$ reweighting procedure is applied in the $n_j = 0$ category. It is estimated by comparing the different effects of reweighting with the nominal weights and with an alternative set of weights derived with a $p_T^{\text{miss}} > 20$ GeV requirement applied in the $Z$-peak region. This requirement follows the event selection criteria used in the $e\mu$ samples where the $Z/\gamma^* \rightarrow \tau\tau$ background contribution is more important. Table XVII shows the uncertainties on the extrapolation factor $\alpha$ to the signal regions and the WW control regions in the $n_j \leq 1$ and $n_j \geq 2$ ggF-enriched categories.

2. $Z/\gamma^* \rightarrow e\mu$ in $n_j \leq 1$

The $f_{\text{recoil}}$ variable (see Sec. IV) shows a clear shape difference between DY and all processes with neutrinos in the final state, including signal and $Z/\gamma^* \rightarrow \tau\tau$, which are collectively referred to as “non-DY”. A method based on a measurement of the selection efficiency of a cut on $f_{\text{recoil}}$ from data, and an estimate of the remaining DY contribution after such a cut, is used in the $e\mu$ category. A sample of events is divided into two bins based on whether they pass or fail the $f_{\text{recoil}}$ requirement, and the former defines the signal region. The efficiency of this cut, $\epsilon = N_{\text{pass}}/(N_{\text{pass}} + N_{\text{fail}})$, measured separately in data for the DY and non-DY processes, is used together with the fraction of the observed events passing the $f_{\text{recoil}}$ requirement to estimate the final DY background. It is analytically equivalent to inverting the matrix:

$$
\begin{bmatrix}
N_{\text{pass}} \\
N_{\text{pass}} + N_{\text{fail}}
\end{bmatrix}
= 
\begin{bmatrix}
1 & 1 \\
1/\epsilon_{\text{DY}} & 1/\epsilon_{\text{non-DY}}
\end{bmatrix}
\begin{bmatrix}
B_{\text{DY}} \\
B_{\text{non-DY}}
\end{bmatrix}
$$

and solving for $B_{\text{DY}}$, which gives the fully data-derived estimate of the DY yield in the $e\mu$ signal region. The $m_T$ distribution for this background is taken from the Monte Carlo prediction, and the $m_T$ shape uncertainties due to the $p_T^{\gamma^*}$ reweighting are found to be negligible.

The non-DY selection efficiency $\epsilon_{\text{non-DY}}$ is evaluated using the $e\mu$ sample, which is almost entirely composed of non-DY events. Since this efficiency is applied to the non-DY events in the final $e\mu$ signal region, the event selection is modified to match the $e\mu$ signal region selection criteria. This efficiency is used for the signal and for all non-DY backgrounds. The DY selection efficiency $\epsilon_{\text{DY}}$ is evaluated using the $e\mu$ sample satisfying the $|m_{\ell\ell} - m_Z| < 15$ GeV requirement, which selects the $Z$-peak region. An additional non-DY efficiency $\epsilon_{\text{non-DY}}'$ is introduced to account for the non-negligible non-DY contribution in the $Z$ peak, and is used in the evaluation of $\epsilon_{\text{DY}}$. It is calculated using the same $m_{\ell\ell}$ region but in $e\mu$ events. Numerical values for these $f_{\text{recoil}}$ selection efficiencies are shown in Table XVIII(a).

---

**TABLE XVII.** $Z/\gamma^* \rightarrow \tau\tau$ uncertainties (in %) on the extrapolation factor $\alpha$, for the $n_j \leq 1$ and $n_j \geq 2$ ggF-enriched categories. Scale, PDF, and generator modeling (Gen) uncertainties are reported. For the $n_j = 0$ category, additional uncertainty due to $p_T^{\gamma^*}$ reweighting is shown. The negative sign indicates anticorrelation with respect to the unsigned uncertainties in the same column.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Scale</th>
<th>PDF</th>
<th>Gen</th>
<th>$p_T^{\gamma^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>-1.6</td>
<td>1.4</td>
<td>5.7</td>
<td>19</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>4.7</td>
<td>1.8</td>
<td>-2.0</td>
<td>-</td>
</tr>
<tr>
<td>$n_j \geq 2$ ggF</td>
<td>-10.3</td>
<td>1.1</td>
<td>10.4</td>
<td>-</td>
</tr>
<tr>
<td>WW control regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>-5.5</td>
<td>1.0</td>
<td>-8.0</td>
<td>16</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>-7.2</td>
<td>2.1</td>
<td>3.2</td>
<td>-</td>
</tr>
</tbody>
</table>

---

**FIG. 27** (color online). $Z/\gamma^* \rightarrow \tau\tau$ control region distributions of transverse mass. See Fig. 5 for plotting details.
TABLE XVIII. The $f_{\text{recoil}}$ summary for the $Z/\gamma^* \rightarrow ee, \mu\mu$ background in the $n_j \leq 1$ categories. The efficiency for Drell-Yan and non-DY processes are given in (a); the associated systematic uncertainties (in %) are given in (b). For each group in (b), the subtotal is given first. The last row gives the total uncertainty on the estimated $B_{\text{DY}}$ yield in the SR.

<table>
<thead>
<tr>
<th>Source</th>
<th>$n_j = 0$</th>
<th>$n_j = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty on $\epsilon_{\text{non-DY}}$</td>
<td>1.9</td>
<td>3.2</td>
</tr>
<tr>
<td>from statistical</td>
<td>1.8</td>
<td>3.0</td>
</tr>
<tr>
<td>from using $ee$ CR to extrapolate to the SR ($ee/\mu\mu$ category)</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>Uncertainty on $\epsilon_{\text{DY}}$</td>
<td>38</td>
<td>32</td>
</tr>
<tr>
<td>from statistical</td>
<td>9.4</td>
<td>16</td>
</tr>
<tr>
<td>from using $Z$ peak to extrapolate to the SR ($12 &lt; m_{ee} &lt; 55$ GeV)</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>Uncertainty on $\epsilon_{\text{non-DY}}$</td>
<td>3.1</td>
<td>4.5</td>
</tr>
<tr>
<td>from statistical</td>
<td>1.9</td>
<td>3.9</td>
</tr>
<tr>
<td>from using $ee$ CR to extrapolate to the SR ($ee/\mu\mu$ category)</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Total uncertainty on yield estimate $B_{\text{DY}}$</td>
<td>49</td>
<td>45</td>
</tr>
</tbody>
</table>

For the non-DY $f_{\text{recoil}}$ selection efficiencies $\epsilon_{\text{non-DY}}$ and $\epsilon_{\text{non-DY}}$, the systematic uncertainties are based on the $ee$-$to-ee/\mu\mu$ extrapolation. They are evaluated with MC simulations by taking the full difference between the selection efficiencies for $ee$ and $ee/\mu\mu$ events in the $Z$ peak and SR. Obtained uncertainties are validated with alternative MC samples and with data, and are added in quadrature to the statistical uncertainties on the efficiencies. The difference in the $f_{\text{recoil}}$ selection efficiencies for the signal and the other non-DY processes is taken as an additional uncertainty on the signal, and is 9% for the $n_j = 0$ category and 7% for the $n_j = 1$ one. Systematic uncertainties on the efficiencies related to the sample composition of the non-DY background processes were found to be negligible.

The systematic uncertainties on $\epsilon_{\text{DY}}$ are based on the extrapolation from the $Z$ peak to the SR and are evaluated with MC simulation by comparing the $f_{\text{recoil}}$ selection efficiencies in these two regions. This procedure is checked with several generators, and the largest difference in the selection efficiency is taken as the systematic uncertainty on the efficiency. It is later added in quadrature to the statistical uncertainty. The procedure is also validated with the data. Table XVIII(b) summarizes all the uncertainties.

For the non-DY $f_{\text{recoil}}$ selection efficiencies $\epsilon_{\text{non-DY}}$ and $\epsilon_{\text{non-DY}}$, the systematic uncertainties are based on the $ee$-$to-ee/\mu\mu$ extrapolation. They are evaluated with MC simulations by taking the full difference between the selection efficiencies for $ee$ and $ee/\mu\mu$ events in the $Z$ peak and SR. Obtained uncertainties are validated with alternative MC samples and with data, and are added in quadrature to the statistical uncertainties on the efficiencies. The difference in the $f_{\text{recoil}}$ selection efficiencies for the signal and the other non-DY processes is taken as an additional uncertainty on the signal, and is 9% for the $n_j = 0$ category and 7% for the $n_j = 1$ one. Systematic uncertainties on the efficiencies related to the sample composition of the non-DY background processes were found to be negligible.

The systematic uncertainties on $\epsilon_{\text{DY}}$ are based on the extrapolation from the $Z$ peak to the SR and are evaluated with MC simulation by comparing the $f_{\text{recoil}}$ selection efficiencies in these two regions. This procedure is checked with several generators, and the largest difference in the selection efficiency is taken as the systematic uncertainty on the efficiency. It is later added in quadrature to the statistical uncertainty. The procedure is also validated with the data. Table XVIII(b) summarizes all the uncertainties.

The largest uncertainties are on $\epsilon_{\text{DY}}$ but since the non-DY component dominates in the signal region, the uncertainties on its $f_{\text{recoil}}$ efficiency are the dominant contribution to the total uncertainty on the estimated $B_{\text{DY}}$ yield.

3. $Z/\gamma^* \rightarrow ee, \mu\mu$ in VBF-enriched $n_j \geq 2$

The $Z/\gamma^* \rightarrow ee, \mu\mu$ background in the VBF-enriched channel is estimated using an “ABCD” method, described below. The BDT shape for this process is taken from a high-purity data sample with low $m_{ee}$ and low $p_T^{\text{miss}}$ (region B). It is then normalized with a $p_T^{\text{miss}}$ cut efficiency, derived from the data using the $Z$-peak region separated into low- and high-$p_T^{\text{miss}}$ regions (C and D, respectively). It yields $0.43 \pm 0.03$. The final estimate in the signal region (A) is corrected with a nonclosure factor derived from the MC, representing the differences in $p_T^{\text{miss}}$ cut efficiencies between the low-$m_{ee}$ and $Z$-peak regions. It yields $0.83 \pm 0.22$. Bins 2 and 3 of $O_{\text{BDT}}$ are normalized using a common factor due to the low number of events in the highest $O_{\text{BDT}}$ bin in region B. The normalization factors, applied to the $Z/\gamma^* \rightarrow ee, \mu\mu$ background in the $ee/\mu\mu$ channel in the signal region, are $\beta_{\text{bin}} = 1.01 \pm 0.15(\text{stat})$ and $\beta_{\text{bin}2+3} = 0.89 \pm 0.28(\text{stat})$.

The uncertainty on the nonclosure factor is 17% (taken as its deviation from unity), and is fully correlated across all $O_{\text{BDT}}$ bins. Uncertainties are included on the $O_{\text{BDT}}$ shape due to QCD scale variations, PDFs, and the parton shower model, and are 11% in the bin with the highest $O_{\text{BDT}}$ score. No dependence of the BDT response on $p_T^{\text{miss}}$ is observed in MC, and an uncertainty is assigned based on the assumption that they are uncorrelated (4%, 10%, and 60% in the bins with increasing $O_{\text{BDT}}$ score).

F. Modifications for 7 TeV data

The background estimation techniques in the $n_j \leq 1$ channels for 7 TeV data closely follow the ones applied to 8 TeV data. The definitions of the control regions of WW, top-quark, and $Z/\gamma^* \rightarrow \tau\tau$ are the same. The $Z/\gamma^* \rightarrow ee, \mu\mu$ background is estimated with the same method based on the $f_{\text{recoil}}$ selection efficiencies. The $f_{\text{recoil}}$ requirements are loosened (see Sec. IV E). The calculation of the extrapolation factor in the $W + \text{jets}$ estimate uses a multijet sample instead of a $Z + \text{jets}$ sample, which has a limited number of events. The VV backgrounds are estimated using Monte Carlo predictions because of the small number of events in the same-charge region. In the $n_j \geq 2$ VBF-enriched category, the background estimation techniques are the same as in the 8 TeV analysis. The normalization factors from the control regions are given in Table XX in the next section along with the values for the 8 TeV analysis.

The theoretical uncertainties on the extrapolation factors used in the WW, top-quark, and $Z/\gamma^* \rightarrow \tau\tau$ background estimation methods are assumed to be the same as in the 8 TeV analysis. Uncertainties due to experimental sources
are unique to the 7 TeV analysis and are taken into account in the likelihood fit. The uncertainties on the $f_{\text{recal}}$ selection efficiencies used in the $Z/\gamma^* \rightarrow ee, \mu\mu$ background estimation were evaluated following the same technique as in the 8 TeV analysis. The dominant uncertainty on the extrapolation factor in the $W + \text{jets}$ estimate is due to the uncertainties on the differences in the compositions of the jets in the multijet and $W + \text{jets}$ sample and is 29% (36%) for muons (electrons).

G. Summary

This section described the control regions used to estimate, from data, the main backgrounds to the various categories in the analysis. An overview of the observed and expected event yields in these control regions is provided in Table XIX for the 8 TeV data. This shows the breakdown of each control region into its targeted physics process (in bold) and its purity, together with the other contributing physics processes. The $WW$ CR in the $n_J = 1$ category is relatively low in $WW$ purity but the normalization for the large contamination by $N_{\text{top}}$ is determined by the relatively pure CR for top quarks.

The normalization factors $\beta$ derived from these control regions are summarized in Table XX, for both the 7 and 8 TeV data samples. Only the statistical uncertainties are quoted and in most of the cases the normalization factors agree with unity within the statistical uncertainties. In two cases where a large disagreement is observed, the systematic uncertainties on $\beta$ are evaluated. One of them is the $WW$ background in the $n_J = 0$ category, where adding the systematic uncertainties reduces the disagreement to about 2 standard deviations: $\beta = 1.22 \pm 0.03(\text{stat}) \pm 0.10(\text{syst})$. The systematic component includes the experimental uncertainties and additionally the theoretical uncertainties on the cross section and acceptance, and the uncertainty on the luminosity determination. Including the systematic uncertainties on the normalization factor for the top-quark background in the first bin in the $n_J \geq 2$ VBF-enriched category reduces the significance of the deviation of the normalization factor with unity: $\beta = 1.58 \pm 0.15(\text{stat}) \pm 0.55(\text{syst})$. In this case, the uncertainty on MC generator modeling is also included. The systematic uncertainties quoted here do not have an impact on the analysis since the background estimation in the signal region is based on the extrapolation factors and their associated uncertainties, as quoted in the previous subsections. In addition, the sample statistics of the control region, the MC sample statistics, and the uncertainties on the background subtraction all affect the estimation of the backgrounds normalized to data.

TABLE XIX. Control region event yields for 8 TeV data. All of the background processes are normalized with the corresponding $\beta$ given in Table XX or with the data-derived methods as described in the text; each row shows the composition of one CR. The $N_{\text{sig}}$ column includes the contributions from all signal production processes. For the VBF-enriched $n_J \geq 2$, the values for the bins in $O_{\text{BDT}}$ are given. The entries that correspond to the target process for the CR are given in bold; this quantity corresponds to $N_{\text{bold}}$ considered in the last column for the purity of the sample (in %). The uncertainties on $N_{\text{bkg}}$ are due to sample size.

<table>
<thead>
<tr>
<th>Control regions</th>
<th>Summary</th>
<th>Composition of $N_{\text{bkg}}$</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_{\text{obs}}$</td>
<td>$N_{\text{bkg}}$</td>
<td>$N_{\text{sig}}$</td>
</tr>
<tr>
<td>$n_J = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR for $WW$</td>
<td>2713</td>
<td>2680 ± 9</td>
<td>28</td>
</tr>
<tr>
<td>CR for top quarks</td>
<td>76013</td>
<td>75730 ± 50</td>
<td>618</td>
</tr>
<tr>
<td>CR for $VV$</td>
<td>533</td>
<td>531 ± 8</td>
<td>2.2</td>
</tr>
<tr>
<td>CR for $Z/\gamma^* \rightarrow \tau\tau$</td>
<td>4557</td>
<td>4530 ± 30</td>
<td>23</td>
</tr>
<tr>
<td>$n_J = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR for $WW$</td>
<td>2647</td>
<td>2640 ± 12</td>
<td>4.3</td>
</tr>
<tr>
<td>CR for top quarks</td>
<td>6722</td>
<td>6680 ± 12</td>
<td>17</td>
</tr>
<tr>
<td>CR for $VV$</td>
<td>194</td>
<td>192 ± 4</td>
<td>1.9</td>
</tr>
<tr>
<td>CR for $Z/\gamma^* \rightarrow \tau\tau$</td>
<td>1540</td>
<td>1520 ± 14</td>
<td>18</td>
</tr>
<tr>
<td>$n_J \geq 2$ ggF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR for top quarks</td>
<td>2664</td>
<td>2660 ± 10</td>
<td>4.9</td>
</tr>
<tr>
<td>CR for $Z/\gamma^* \rightarrow \tau\tau$</td>
<td>266</td>
<td>263 ± 6</td>
<td>2.6</td>
</tr>
<tr>
<td>$n_J \geq 2$ VBF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR for top quarks, bin 1</td>
<td>143</td>
<td>142 ± 2</td>
<td>2.1</td>
</tr>
<tr>
<td>CR for top quarks, bin 2–3</td>
<td>14</td>
<td>14.3 ± 0.5</td>
<td>1.8</td>
</tr>
<tr>
<td>CR for $Z/\gamma^* \rightarrow \tau\tau$</td>
<td>24</td>
<td>20.7 ± 0.9</td>
<td>2.4</td>
</tr>
</tbody>
</table>
VII. FIT PROCEDURE AND UNCERTAINTIES

The signal yields and cross sections are obtained from a statistical analysis of the data samples described in Sec. IV. A likelihood function—defined to simultaneously model, or “fit” the yields of the various subsamples—is maximized. The signal strength parameter $\mu$, defined in Sec. I, is the ratio of the measured signal yield to the expected SM value. Its expected value ($\mu_{\exp}$) is unity by definition. A measurement of zero corresponds to no signal in the data. The observed value $\mu_{\text{obs}}$, reported in Sec. IX, is one of the central results of this paper.

In this section, the fit regions are described in Sec. VII A followed by the details of the likelihood function and the test statistic in Sec. VII B. Section VII C summarizes the various sources of uncertainty that affect the results. A check of the results is given in Sec. VII D.

A. Fit regions

The fit is performed over data samples defined by fit regions listed in Table XXI, which consist of
(i) signal region categories [Table XXI(a)] and
(ii) profiled control regions [rows in Table XXI(b) marked by solid circles].

The nonprofiled control regions [rows in Table XXI(b) marked by open circles] do not have explicit terms in the likelihood, but are listed in the table for completeness.

The profiled CRs determine the normalization of the corresponding backgrounds through a Poisson term in the likelihood, which, apart from the Drell-Yan $\tau\tau$ CR, use the $ee/\mu\mu$ sample. The nonprofiled CRs do not have a Poisson term and enter the fit in other ways. The details are described in the next section.

The SR categories $i$ and fit distribution bins $b$ that contribute to the likelihood were briefly motivated in Sec. II.

The $ee/\mu\mu$ samples in the $n_j \leq 1$, the most signal sensitive of all channels, are each divided into 12 kinematic regions ($12 = 2 \cdot 3 \cdot 2$): two regions in $m_{ll}$, three regions in $p_T^{l2}$, and two regions for the subleading lepton flavors. In contrast, the less sensitive $ee/\mu\mu$ samples for the $n_j \leq 1$ categories use one range of $m_{ll}$ and $p_T^{l2}$.

The $m_T$ distribution is used to fit all of the ggF-enriched categories. Its distribution for the signal process has an upper kinematic edge to $m_H$ but, in practice, $m_T$ can exceed $m_H$ because of detector resolution. There is also a kinematic suppression below a value of $m_T$ that increases with increasing values of $m_{ll}$ and $p_T^{l2}$ due to the kinematic requirements in each of the $n_j \leq 1$ categories.

The $m_T$ distribution for the $n_j = 0$ category in the $ee/\mu\mu$ samples uses a variable binning scheme that is optimized for each of the 12 (one) kinematic regions. In the kinematically favored range of the $ee/\mu\mu$ samples, there are 8 bins that are approximately 6 GeV wide between a range of $x$ to $y$, where $x$ is approximately 80 GeV and $y$ is approximately 130 GeV. A single bin at low $m_T$, from 0 to $x$, has a few events in each category; another bin at high $m_T$—from $y$ to $\infty$—is populated dominantly by $WW$ and top-quark events, constraining these backgrounds in the fit.

The $m_T$ distribution for the $n_j = 1$ category follows the above scheme with six bins. The bins are approximately 10 GeV wide in the same range as for $n_j = 0$.

The $m_T$ distribution of the $ee/\mu\mu$ events in the ggF-enriched $n_j \geq 2$ uses four bins specified by the bin boundaries $[0, 50, 80, 130, \infty]$ GeV.

The $O_{\text{BDT}}$ distribution is used to fit the VBF-enriched $n_j \geq 2$ samples. The signal purity increases with increasing value of $O_{\text{BDT}}$, so the bin widths decrease accordingly. The bin boundaries are $[-1, -0.48, 0.3, 0.78, 1]$ and define four bins that are labeled 0 through 3. Only bins 1, 2, and 3 are used in the fit. The selection-based cross-check analysis uses two bins in $m_{j1}([600, 1000, \infty]$ GeV) and four bins in the $m_T$ distribution as defined above for the ggF-enriched $n_j \geq 2$. 

### Table XX. Control region normalization factors $\beta$. The $\beta$ values scale the corresponding estimated yields in the signal region; those that use MC-based normalization are marked with a dash. For the VBF-enriched $n_j \geq 2$ category, the values in bins of $O_{\text{BDT}}$ are given for top quarks; a combined value is given for $Z/\gamma^* \rightarrow \tau\tau$. The uncertainties are due to the sample size of the corresponding control regions.

<table>
<thead>
<tr>
<th>Category</th>
<th>WW</th>
<th>Top quarks</th>
<th>$VV$</th>
<th>$Z/\gamma^* \rightarrow \tau\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 TeV sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>1.22 ± 0.03</td>
<td>1.08 ± 0.02</td>
<td>0.92 ± 0.07</td>
<td>1.00 ± 0.02</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>1.05 ± 0.05</td>
<td>1.06 ± 0.03</td>
<td>0.96 ± 0.12</td>
<td>1.05 ± 0.04</td>
</tr>
<tr>
<td>$n_j \geq 2$, ggF</td>
<td>-</td>
<td>1.05 ± 0.03</td>
<td>-</td>
<td>1.00 ± 0.09</td>
</tr>
<tr>
<td>$n_j \geq 2$, VBF bin 1</td>
<td>-</td>
<td>1.58 ± 0.15</td>
<td>-</td>
<td>0.90 ± 0.30</td>
</tr>
<tr>
<td>$n_j \geq 2$, VBF bins 2–3</td>
<td>-</td>
<td>0.95 ± 0.31</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7 TeV sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 0$</td>
<td>1.09 ± 0.08</td>
<td>1.12 ± 0.06</td>
<td>-</td>
<td>0.89 ± 0.04</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>0.98 ± 0.12</td>
<td>0.99 ± 0.04</td>
<td>-</td>
<td>1.10 ± 0.09</td>
</tr>
<tr>
<td>$n_j \geq 2$, VBF bins 1–3</td>
<td>-</td>
<td>0.82 ± 0.29</td>
<td>-</td>
<td>1.52 ± 0.91</td>
</tr>
</tbody>
</table>
TABLE XXI. Fit region definitions for the Poisson terms in the likelihood, Eq. (11), not including the terms used for MC statistics. The signal region categories $i$ are given in (a). The definitions for bins $b$ are given by listing the bin edges, except for $m_T$ and $O_{\text{BDT}}$, and are given in the text and noted as the fit variables on the right-most column. The background control regions are given in (b), which correspond to the ones indicated as using data in Table XI. The profiled CRs are marked by $\ast$ and the others are marked by $\circ$. “Sample” notes the lepton flavor composition of the CR that is used for all the SR regions for a given $n_j$ category: “$e\mu$” means that a $e\mu$ CR sample is used for all SR regions; the $Wj$ and $jj$ CRs use the same lepton-flavor samples in the SR (same), i.e., “$e\mu$” CR for “$e\mu$” SR and “$ee/\mu\mu$” CR for “$ee/\mu\mu$” SR; the DY, $ee/\mu\mu$ sample is used only for the $ee/\mu\mu$ SR; the two rows in $n_j \geq 2$ VBF use a CR that combines the two samples (both); see text for details. Energy-related quantities are in GeV.

<table>
<thead>
<tr>
<th>$n_j$, flavor</th>
<th>$m_{T_{\ell \ell}}$</th>
<th>$p_T^{T_{\ell \ell}}$</th>
<th>$\ell_2$</th>
<th>Fit variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_j = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e\mu$</td>
<td>$[10, 30, 55]$</td>
<td>$[10, 15, 20]$</td>
<td>$[e, \mu]$</td>
<td>$m_T$</td>
</tr>
<tr>
<td>$ee/\mu\mu$</td>
<td>$[12, 55]$</td>
<td>$[10, \infty]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e\mu$</td>
<td>$[10, 30, 55]$</td>
<td>$[10, 15, 20, \infty]$</td>
<td>$[e, \mu]$</td>
<td>$m_T$</td>
</tr>
<tr>
<td>$ee/\mu\mu$</td>
<td>$[12, 55]$</td>
<td>$[10, \infty]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j \geq 2$ ggF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e\mu$</td>
<td>$[10, 55]$</td>
<td>$[10, \infty]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_j \geq 2$ VBF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e\mu$</td>
<td>$[10, 50]$</td>
<td>$[10, \infty]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ee/\mu\mu$</td>
<td>$[12, 50]$</td>
<td>$[10, \infty]$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CR</th>
<th>Profiled?</th>
<th>Sample</th>
<th>Notable differences vs. SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_j = 0$</td>
<td></td>
<td>$e\mu$</td>
<td>$55 &lt; m_{T_{\ell \ell}} &lt; 110, \Delta \phi_{T_{\ell \ell}} &lt; 2.6, p_T^{T_{\ell \ell}} &gt; 15$</td>
</tr>
<tr>
<td>WW</td>
<td>$\ast$</td>
<td>$e\mu$</td>
<td>$n_j = 0$ after presel., $\Delta \phi_{T_{\ell \ell}} &lt; 2.8$</td>
</tr>
<tr>
<td>Top</td>
<td>$\circ$</td>
<td>$e\mu$</td>
<td>one anti-identified $\ell$</td>
</tr>
<tr>
<td>Wj</td>
<td>$\circ$</td>
<td>same</td>
<td>two anti-identified $\ell$</td>
</tr>
<tr>
<td>jj</td>
<td>$\circ$</td>
<td>same</td>
<td>same-charge $\ell$ (only used in $e\mu$)</td>
</tr>
<tr>
<td>VV</td>
<td>$\ast$</td>
<td>$e\mu$</td>
<td>$f_{\text{recoil}} &gt; 0.1$ (only used in $ee/\mu\mu$)</td>
</tr>
<tr>
<td>DY, $ee/\mu\mu$</td>
<td>$\ast$</td>
<td>$ee/\mu\mu$</td>
<td>$m_{T_{\ell \ell}} &lt; 80, \Delta \phi_{T_{\ell \ell}} &gt; 2.8$</td>
</tr>
</tbody>
</table>

| $n_j = 1$ |            | $e\mu$ | $m_{T_{\ell \ell}} > 80, |m_{T_{\ell \ell}} - m_Z| > 25, p_T^{T_{\ell \ell}} > 15$ |
| WW | $\ast$ | $e\mu$ | $n_b = 1$ |
| Top | $\circ$ | $e\mu$ | one anti-identified $\ell$ |
| Wj | $\circ$ | same | two anti-identified $\ell$ |
| jj | $\circ$ | same | same-charge $\ell$ (only used in $e\mu$) |
| VV | $\ast$ | $e\mu$ | $f_{\text{recoil}} > 0.1$ (only used in $ee/\mu\mu$) |
| DY, $ee/\mu\mu$ | $\ast$ | $ee/\mu\mu$ | $m_{T_{\ell \ell}} < 80, m_{T_{\ell \ell}} > m_Z - 25$ |

| $n_j \geq 2$ ggF | $\ast$ | $e\mu$ | $m_{T_{\ell \ell}} > 80$ |
| Top | $\circ$ | $e\mu$ | one anti-identified $\ell$ |
| Wj | $\circ$ | same | two anti-identified $\ell$ |
| jj | $\circ$ | same | same-charge $\ell$ (only used in $e\mu$) |
| DY, $ee/\mu\mu$ | $\ast$ | $ee/\mu\mu$ | $m_{T_{\ell \ell}} < 80, |m_{T_{\ell \ell}} - m_Z| > 25$ |

| $n_j \geq 2$ VBF | $\ast$ | both | $n_b = 1$ |
| Top | $\circ$ | same | one anti-identified $\ell$ |
| Wj | $\circ$ | same | two anti-identified $\ell$ |
| Dij | $\ast$ | $ee/\mu\mu$ | $E^{\text{miss}} < 45$ (only used in $ee/\mu\mu$) |
| DY, $e\mu$ | $\circ$ | both | $m_{T_{\ell \ell}} < 80, |m_{T_{\ell \ell}} - m_Z| < 25$ |
In general, the bin boundaries are chosen to maximize the expected signal significance while stabilizing the statistical fluctuations associated with the subtraction of the backgrounds. For the $m_T$ fits, this is accomplished by maintaining an approximately constant signal yield in each of the bins. The exact values of the $m_T$ bins are given in Appendix A1 in Table XXIX. 

The interplay of the various fit regions is illustrated for one kinematic region of the $n_j = 0$ in Fig. 28. The shape of the $m_T$ distribution is used in the fit to discriminate between the signal and the background as shown in the top row for the SR. Three profiled CRs determine the normalization factors ($\beta_k$) of the respective background contributions in situ. The variable and selections used to separate the SR from the CRs regions are given in the second row: for $WW$ the $m_{ll}$ variable divides the SR and CR, but also the validation region (VR) used to test the $WW$ extrapolation (see Sec. VI A); for $DY$ the $\Delta \phi_{ll}$ variable divides the SR and CR with a region separating the two; for $VV$ the discrete same/opposite charge variable is used. The last row shows the backgrounds whose normalizations are not profiled in the fit, but are computed prior to the fit.

The treatment of a given region as profiled or non-profiled CR depends from the complexity related to its

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig28.png}
\caption{(color online) Simplified illustration of the fit regions for $n_j = 0$, $e\mu$ category. The figure in (a) is the variable-binned $m_T$ distribution in the signal region for a particular range of $m_{ll}$ and $p_T$ specified in Table XXI; the $m_T$ bins are labeled $b = 1, 2, \ldots$; the histograms are stacked for the five principal background processes—$WW$, top, Misid (mostly $Wj$), $VV$, $DY$ (unlabeled)—and the Higgs signal process. The figures in (b, c, d) represent the distributions that define the various profiled control regions used in the fit with a corresponding Poisson term in the likelihood $\mathcal{L}$. Those in (e, f, g) represent the nonprofiled control regions that do not have a Poisson term in $\mathcal{L}$, but determine parameters that modify the background yield predictions. A validation region (VR) is also defined in (b); see text.}
\end{figure}
implementation in the fit, the impact of the estimated background in the analysis, and the level of contamination of the other process in the relative CR. Subdominant backgrounds and those whose estimation is not largely affected by the postfit yield of the other backgrounds, like $Wj$ and multijet backgrounds, are not profiled.

B. Likelihood, exclusion, and significance

The statistical analysis involves the use of the likelihood $\mathcal{L}(\mu, \theta | N)$, which is a function of the signal strength parameter $\mu$ and a set of nuisance parameters $\theta = \{\theta_1, \theta_2, \ldots\}$ given a set of the numbers of events $N = \{N_A, N_B, \ldots\}$. Allowed ranges of $\mu$ are found using the distribution of a test statistic $q_{\mu}$.

\[
\mathcal{L} = \prod_{b,i} f\left(N_{ib} \mid \mu \cdot S_{ib} \prod_r v_{br}(\theta_r) + \sum_k \beta_k \cdot B_{kib} \prod_s v_{bs}(\theta_s) \right) + \prod_{l} \prod_{k} f\left(N_{li} \mid \sum_k \beta_k \cdot B_{kl} \prod_t g(\theta_t | \theta_{t'}) \right) + \prod_{k} f\left(\xi_k \mid \zeta \cdot \theta_k \right).
\]

The statistical uncertainties are considered explicitly in the first, second, and fourth terms. The first and second terms treat the random error associated with the predicted value, i.e., for a background yield estimate $B$ the $\sqrt{B}$ error associated with it. The fourth term treats the sampling error associated with the finite sample size used for the prediction, e.g., the $\sqrt{N_{MC}}$ "MC statistical errors" when MC is used. All of the terms are described below and summarized in Eq. (11).

The first term of $\mathcal{L}$ is a Poisson function $f$ for the probability of observing $N$ events given $\lambda$ expected events, $f(N \mid \lambda) = e^{-\lambda} \lambda^N / N!$. The expected value $\lambda$ is the sum of event yields from signal $(S)$ and the sum of the background contributions $(\Sigma_k B_k)$ in a given signal region, i.e., $\lambda = \mu \cdot S + \Sigma_k B_k$. The parameter of interest, $\mu$, multiplies $S$; each background yield in the sum is evaluated as described in Sec. VI. In our notation, the yields are scaled by the response functions $\nu$ that parametrize the impact of the systematic uncertainties $\theta$. The $\nu$ and $\theta$ are described in more detail below when discussing the third term of $\mathcal{L}$.

The second term constrains the background yields with Poisson components that describe the profiled control regions. Each term is of the form $f(N_{li} \mid \lambda_i)$ for a given CR labeled by $l$, where $N_{li}$ is the number of observed events in $l$, i.e., $\lambda_i = \sum_s \beta_s \cdot B_{kl}$ is the predicted yield in $l$, $\beta_s$ is the normalization factor of background $k$, and $B_{kl}$ is the MC or data-derived estimate of background $k$ in $l$. The $\beta_s$ parameters are the same as those that appear in the first Poisson component above.

The third term constrains the systematic uncertainties with Gaussian terms. Each term is of the form $g(\theta \mid \delta) = e^{-\left(\theta - \delta\right)^2 / 2} / \sqrt{2\pi}$, where $\theta$ represents the central value of the measurement and $\delta$ the associated nuisance parameter for a given systematic uncertainty. The effect of the systematic uncertainty on the yields is through an exponential response function $\nu(\theta) = (1 + \epsilon)\theta$ for normalization uncertainties that have no variations among bins $b$ of the fit variable, where $\epsilon$ is the value of the uncertainty in question. In this case, $\nu$ follows a log-normal distribution [96]. In this notation, $\epsilon = 3\%$ is written if the uncertainty that corresponds to one standard deviation affects the associated yield by $\pm 3\%$ and corresponds to $\theta = \pm 1$, respectively.

For the cases where the systematic uncertainty affects a given distribution differently in each bin $b$, a different linear response function is used in each bin; this function is written as $\nu_b(\theta) = 1 + \epsilon_b \cdot \theta$. In this case, $\nu_b$ is normally distributed around unity with width $\epsilon_b$, and is truncated by the $\nu_b > 0$ restriction to avoid unphysical values. Both types of response function impact the predicted $S$ and $B_k$ in the first Poisson component.

The fourth term treats the sample error due to the finite sample size [97], e.g., the sum of the number of generated MC events for all background processes, $B = \Sigma_k B_k$. The quantity $B$ is constrained with a Poisson term $f(\xi \mid \lambda)$, where $\xi$ represents the central value of the background estimate and $\lambda = \xi \cdot \theta$. The $\zeta = (B / \lambda)^2$ defines the quantity with the statistical uncertainty of $B$ as $\delta$. For instance, if a
background yield estimate \( B \) uses \( N_{\text{MC}} \) MC events that correspond to a data sample with effective luminosity \( L_{\text{MC}} \), then for a data-to-MC luminosity ratio \( r = L_{\text{data}}/L_{\text{MC}} \) the background estimate is \( B = r \cdot N_{\text{MC}} \), and the uncertainty parameter in question is \( \delta = r \cdot \sqrt{N_{\text{MC}}}, \) \( \zeta = N_{\text{MC}} \). In this example, the Poisson function is evaluated at \( N_{\text{MC}} \) given \( \lambda = \theta \cdot N_{\text{MC}} \). Similar to the case for the third term, a linear response function \( \nu(\theta) = \theta \) impacts the predicted \( S \) and \( B_k \) in the first Poisson component.

In summary, the likelihood is the product of the four above-mentioned terms and can be written schematically as done in Eq. (11), where the \( \nu_{br} \) and \( \nu_{bs} \) are implicitly products over all three types of response functions—normalization, shape of the distribution, and finite MC sample size—whose parameters are constrained by the second, third, and fourth terms, respectively. In the case of finite MC sample size, \( \theta \) is unique to each bin, which is not shown in Eq. (11). The statistical treatments of two quantities—\( \gamma' \rightarrow ee \), \( \mu \mu \) estimate in \( n_j \leq 1 \) and the top-quark estimate in \( n_j = 1 \)—are constrained with additional multiplicative terms in the likelihood (see Appendix A).

To determine the observed value of the signal strength, \( \mu_{\text{obs}} \), the likelihood is maximized with respect to its arguments, \( \mu \) and \( \theta \), and evaluated at \( \theta = 0 \) and \( \xi = \zeta \).

### 2. Test statistic

The profiled likelihood-ratio test statistic [98] is used to test the background-only or background-and-signal hypotheses. It is defined as

\[
q(\mu) = -2 \ln \frac{\mathcal{L}(\mu, \theta)}{\mathcal{L}_{\text{max}} \mid \theta = \hat{\theta}_r},
\]

and it is also written as \( q_0 \), the argument of the logarithm is written as \( \Lambda \) in later plots. The denominator of Eq. (12) is unconditionally maximized over all possible values of \( \mu \) and \( \theta \), while the numerator is maximized over \( \theta \) for a conditional value of \( \mu \). The latter takes the values \( \hat{\theta}_r \), which are \( \theta \) values that maximize \( \mathcal{L} \) for a given value of \( \mu \). When the denominator is maximized, \( \mu \) takes the value of \( \bar{\mu} \).

The \( p_0 \) value is computed for the test statistic \( q_0 \), Eq. (12), evaluated at \( \mu = 0 \), and is defined to be the probability to obtain a value of \( p_0 \) larger than the observed value under the background-only hypothesis. There are no boundaries on \( \bar{\mu} \), although \( q_0 \) is defined to be negative if \( \bar{\mu} \leq 0 \). All \( p_0 \) values are computed using the asymptotic approximation that \( -2 \ln \Lambda(\mu) \) follows a \( \chi^2 \) distribution [98].

Local significance is defined as the one-sided tail of a Gaussian distribution, \( Z_0 = \sqrt{2 \ln \Lambda^{-1}(1 - 2 p_0)} \).

A modified frequentist method known as CL\(_S\) [99] is used to compute the one-sided 95% confidence level (C.L.) exclusion regions.

### 3. Combined fit

The combined results for the 7 and 8 TeV data samples account for the correlations between the analyses due to common systematic uncertainties.

The correlation of all respective nuisance parameters is assumed to be 100% except for those that are statistical in origin or have a different source for the two data sets. Uncorrelated systematics include the statistical component of the jet energy scale calibration and the luminosity uncertainty. All theoretical uncertainties are treated as correlated.

### C. Systematic uncertainties

Uncertainties enter the fit as nuisance parameters in the likelihood function [Eq. (11)]. Uncertainties (both theoretical and experimental) specific to individual processes are described in Secs. V and VI; experimental uncertainties common to signal and background processes are described in this subsection. The impact on the yields and distributions from both sources of uncertainty is also discussed.

#### 1. Sources of uncertainty

The dominant sources of experimental uncertainty on the signal and background yields are the jet energy scale and resolution, and the \( b \)-tagging efficiency. Other sources of uncertainty are the lepton resolutions and identification and trigger efficiencies, missing transverse momentum measurement, and the luminosity calculation. The uncertainty on the integrated luminosity in the 8 TeV data analysis is 2.8%. It is derived following the same methodology as in Ref. [100], from a preliminary calibration of the luminosity scale derived from beam-separation scans. The corresponding uncertainty in the 7 TeV data analysis is 1.8%.

The jet energy scale is determined from a combination of test beam, simulation, and \( \textit{in situ} \) measurements [27]. Its uncertainty is split into several independent categories: modeling and statistical uncertainties on the extrapolation of the jet calibration from the central region (\( \eta \) intercalibration), high-\( p_T \) jet behavior, MC nonclosure uncertainties, uncertainties on the calorimeter response and calibration of the jets originating from light quarks or gluons, the \( b \)-jet energy scale uncertainties, uncertainties due to modeling of in-time and out-of-time pile-up, and uncertainties on \( \textit{in situ} \) jet energy corrections. All of these categories are further subdivided by the physical source of the uncertainty. For jets used in this analysis (\( p_T > 25 \) GeV and \( |\eta| \leq 4.5 \)), the jet energy scale uncertainty ranges from 1% to 7%, depending on \( p_T \) and \( \eta \). The jet energy resolution varies from 5% to 20% as a function of the jet \( p_T \) and \( \eta \). The relative uncertainty on the resolution, as determined from \( \textit{in situ} \) measurements, ranges from 2% to 40%, with the largest value of the resolution and relative uncertainty occurring at the \( p_T \) threshold of the jet selection.
The method used to evaluate the $b$-jet tagging efficiency uses a sample dominated by dileptonic decays of top-quark pairs. This method is based on a likelihood fit to the data, which combines the per-event jet-flavor information and the expected momentum correlation between the jets to allow the $b$-jet tagging efficiency to be measured to high precision [30]. To further improve the precision, this method is combined with a second calibration method, which is based on samples containing muons reconstructed in the vicinity of the jet. The uncertainties related to $b$-jet identification are decomposed into six uncorrelated components using an eigenvector method [32]. The number of components is equal to the numbers of $p_T$ bins used in the calibration, and the uncertainties range from < 1% to 7.8%. The uncertainties on the misidentification rate for light-quark jets depend on $p_T$ and $\eta$, and have a range of 9%–19%. The uncertainties on $c$-jets reconstructed as $b$-jets range between 6% and 14% depending on $p_T$ only.

The reconstruction, identification, isolation, and trigger efficiencies for electrons and muons, as well as their momentum scales and resolutions, are estimated using $Z \to ee/\mu\mu$, $J/\psi \to ee/\mu\mu$, and $W \to e\nu/\mu\nu$ decays [20, 23]. The uncertainties on the lepton and trigger efficiencies are smaller than 1% except for the uncertainty on the electron identification efficiency, which varies between 0.2% and 2.7% depending on the electron identification efficiency, which is based on samples containing muons reconstructed in the vicinity of the jet. The uncertainties related to $b$-jet identification are decomposed into six uncorrelated components using an eigenvector method [32]. The number of components is equal to the numbers of $p_T$ bins used in the calibration, and the uncertainties range from < 1% to 7.8%. The uncertainties on the misidentification rate for light-quark jets depend on $p_T$ and $\eta$, and have a range of 9%–19%. The uncertainties on $c$-jets reconstructed as $b$-jets range between 6% and 14% depending on $p_T$ only.

The reconstruction, identification, isolation, and trigger efficiencies for electrons and muons, as well as their momentum scales and resolutions, are estimated using $Z \to ee/\mu\mu$, $J/\psi \to ee/\mu\mu$, and $W \to e\nu/\mu\nu$ decays [20, 23]. The uncertainties on the lepton and trigger efficiencies are smaller than 1% except for the uncertainty on the electron identification efficiency, which varies between 0.2% and 2.7% depending on $p_T$ and $\eta$, and the uncertainties on the isolation efficiencies, which are the largest for $p_T < 15$ GeV and yield 1.6% and 2.7% for electrons and muons, respectively.

The changes in jet energy and lepton momenta due to varying them by their systematic uncertainties are propagated to $E_T^{\text{miss}}$; the changes in the high-$p_T$ object momenta and in $E_T^{\text{miss}}$ are, therefore, fully correlated [33]. Additional contributions to the $E_T^{\text{miss}}$ uncertainty arise from the modeling of low-energy particle measurements (soft terms). In the calorimeter, these particles are measured as calibrated clusters of cells that are above a noise threshold but not associated with reconstructed physics objects. The longitudinal and perpendicular (with respect to the hard component of the missing transverse momentum) components of the soft terms are fit with Gaussian functions in data and MC DY samples in order to assess the associated uncertainties. The uncertainties are parameterized as a function of the magnitude of the summed $p_T$ of the high-$p_T$ objects, and are evaluated in bins of the average number of interactions per bunch crossing. Differences of the mean and width of the soft term components between data and MC result in variations on the mean of the longitudinal component of about 0.2 GeV, while the uncertainty on the resolution of the longitudinal and perpendicular components is 2% on average.

Jet energy and lepton momentum scale uncertainties are also propagated to the $p_T^{\text{miss}}$ calculation. The systematic uncertainties related to the track-based soft term are based on the balance between tracks not associated with charged leptons and jets and the total transverse momentum of the hard objects in the event. These uncertainties are calculated by comparing the properties of $p_T^{\text{miss}}$ in $Z \to ee/\mu\mu$ events in real and simulated data, as a function of the magnitude of the summed $p_T$ of the hard $p_T$ objects in the event. The variations on the mean of the longitudinal component are in the range 0.3–1.4 GeV and the uncertainties on the resolution on the longitudinal and perpendicular components are in the range 1.5–3.3 GeV, where the lower and upper bounds correspond to the range of the sum of the hard $p_T$ objects below 5 GeV and above 50 GeV, respectively.

2. Impact on yields and distributions

In the likelihood fit, the experimental uncertainties are varied in a correlated way across all backgrounds and all signal and control regions, so that uncertainties on the extrapolation factors $\alpha$ described in Sec. VI are correctly propagated. If the normalization uncertainties are less than 0.1% they are excluded from the fit. If the shape uncertainties (discussed below) are less than 1% in all bins, they are excluded as well. Removing such small uncertainties increases the performance and stability of the fit.

In the fit to the $m_T$ distribution to extract the signal yield, the predicted $m_T$ shape from simulation is used for all of the backgrounds except $W +$ jets and multijets. The impact of experimental uncertainties on the $m_T$ shapes for the individual backgrounds and signal are evaluated, and no significant impact is observed for the majority of the experimental uncertainties. Those experimental uncertainties that do produce statistically significant variations of the shape have no appreciable effect on the final results, because the uncertainty on the $m_T$ shape of the total background is dominated by the uncertainties on the normalizations of the individual backgrounds. The theoretical uncertainties on the $WW$ and $Wy$ $m_T$ shape are considered in the $nj \leq 1$ categories, as discussed in Secs. VI A 1 and VI D. In the $nj \geq 2$ ggF-enriched category, only the theoretical uncertainties on the top-quark $m_T$ shape are included (see Sec. VI B 4).

The $O_{\text{BDT}}$ output distribution is fit in the $nj \geq 2$ VBF-enriched category, and as with the $m_T$ distribution its shape is taken from the MC simulation, except for the $W +$ jets and multijet background processes. The theoretical uncertainties on the top-quark $O_{\text{BDT}}$ shape are included in the analysis, as described in Sec. VI B 4.

Table XXII(a) shows the relative uncertainties on the combined predicted signal yield, summed over all the lepton-flavor channels, for each $nj$ category for the 8 TeV analysis. They represent the final postfit uncertainties on the estimated yields. The first two entries show the perturbative uncertainties on the ggF-jet-bin acceptances in the exclusive $nj = 0$ and $nj = 1$ categories. The following entries are specific to the QCD scale uncertainties on the inclusive $nj \geq 2$ and $nj \geq 3$ cross sections, and on the total
cross section and the acceptance. The latter includes the uncertainties due to the PDF variations, UE/PS and generator modeling, as described in Table X. The uncertainties on the VBF production process are also shown but are of less importance. The dominant uncertainties on the signal yields are theoretical. The uncertainties on the \( f_{\text{recoil}} \) selection efficiency (relevant to the \( Z/\gamma^* \rightarrow e\mu \) estimate in the \( n_j \leq 1 \) categories) are applied only in the \( ee/\mu\mu \) channels.

Table XXII(b) shows the leading uncertainties on the cumulative background yields for each \( n_j \) category. The first three entries are theoretical and apply to the WW, top-quark, and \( VV \) processes (see Sec. VI). The remaining uncertainties arise from the modeling of specific backgrounds and from experimental uncertainties.

Table XXIII summarizes the above postfit uncertainties on the total signal and backgrounds yields. The uncertainties shown are divided into three categories: statistical, experimental, and theoretical. The statistical uncertainties are only relevant in the cases where the background estimates rely on the data. For example, the entry under \( N_{WW} \) in \( n_j = 0 \) represents the uncertainty on the sample statistics in the WW control region. The uncertainties on \( N_{\text{top}} \) in the \( n_j \leq 1 \) categories also include the uncertainties on the corrections applied to the normalization factors. The uncertainties from the number of events in the control samples used to derive the \( W + \text{jets} \) and multijet extrapolation factors are listed under the experimental category, as discussed in Sec. VI C. Uncertainties on the total \( W + \text{jets} \) estimate are reduced compared to the values quoted in

<table>
<thead>
<tr>
<th>Source of Uncertainty</th>
<th>( n_j = 0 )</th>
<th>( n_j = 1 )</th>
<th>( n_j \geq 2 ) ( ggF )</th>
<th>( n_j \geq 2 ) VBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ggF ) H, jet veto for ( n_j = 0 ), ( \epsilon_0 )</td>
<td>8.1</td>
<td>14</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>( ggF ) H, jet veto for ( n_j = 1 ), ( \epsilon_1 )</td>
<td>-</td>
<td>12</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>( ggF ) H, ( n_j \geq 2 ) cross section</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.9</td>
</tr>
<tr>
<td>( ggF ) H, ( n_j \geq 3 ) cross section</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.1</td>
</tr>
<tr>
<td>( ggF ) H, total cross section</td>
<td>10</td>
<td>9.1</td>
<td>7.9</td>
<td>2.0</td>
</tr>
<tr>
<td>( ggF ) H acceptance model</td>
<td>4.8</td>
<td>4.5</td>
<td>4.2</td>
<td>4.0</td>
</tr>
<tr>
<td>VBF H, total cross section</td>
<td>-</td>
<td>0.4</td>
<td>0.8</td>
<td>2.9</td>
</tr>
<tr>
<td>VBF H acceptance model</td>
<td>-</td>
<td>0.3</td>
<td>0.6</td>
<td>5.5</td>
</tr>
<tr>
<td>( H \rightarrow WW^* ) branching fraction</td>
<td>4.3</td>
<td>4.3</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Jet energy scale &amp; resolution</td>
<td>5.1</td>
<td>2.3</td>
<td>7.1</td>
<td>5.4</td>
</tr>
<tr>
<td>( p_T^{\text{miss}} ) scale &amp; resolution</td>
<td>0.6</td>
<td>1.4</td>
<td>0.1</td>
<td>1.2</td>
</tr>
<tr>
<td>( f_{\text{recoil}} ) efficiency</td>
<td>2.5</td>
<td>2.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trigger efficiency</td>
<td>0.8</td>
<td>0.7</td>
<td>-</td>
<td>0.4</td>
</tr>
<tr>
<td>Electron id., isolation, reconstruction eff.</td>
<td>1.4</td>
<td>1.6</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Muon id., isolation, reconstruction eff.</td>
<td>1.1</td>
<td>1.6</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Pile-up model</td>
<td>1.2</td>
<td>0.8</td>
<td>0.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Uncertainty</th>
<th>( N_{\text{bkg}} ) (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW theoretical model</td>
<td>1.4</td>
</tr>
<tr>
<td>Top theoretical model</td>
<td>-</td>
</tr>
<tr>
<td>VV theoretical model</td>
<td>-</td>
</tr>
<tr>
<td>( Z/\gamma^* \rightarrow \tau\tau ) estimate</td>
<td>0.6</td>
</tr>
<tr>
<td>( Z/\gamma^* \rightarrow ee, \mu\mu ) estimate in VBF</td>
<td>-</td>
</tr>
<tr>
<td>( Wj ) estimate</td>
<td>1.0</td>
</tr>
<tr>
<td>( jj ) estimate</td>
<td>0.1</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>-</td>
</tr>
<tr>
<td>Jet energy scale &amp; resolution</td>
<td>0.4</td>
</tr>
<tr>
<td>( p_T^{\text{miss}} ) scale &amp; resolution</td>
<td>0.1</td>
</tr>
<tr>
<td>( b )-tagging efficiency</td>
<td>-</td>
</tr>
<tr>
<td>Light- and c-jet mistag</td>
<td>-</td>
</tr>
<tr>
<td>( f_{\text{recoil}} ) efficiency</td>
<td>0.5</td>
</tr>
<tr>
<td>Trigger efficiency</td>
<td>0.3</td>
</tr>
<tr>
<td>Electron id., isolation, reconstruction eff.</td>
<td>0.3</td>
</tr>
<tr>
<td>Muon id., isolation, reconstruction eff.</td>
<td>0.2</td>
</tr>
<tr>
<td>Pile-up model</td>
<td>0.4</td>
</tr>
</tbody>
</table>
TABLE XXIII. Composition of the postfit uncertainties (in %) on the total signal (N_{SIG}), background (N_{BKG}), and individual background yields in the signal regions. The total uncertainty (Total) is decomposed into three components: statistical (Stat), experimental (Expt), and theoretical (Theo). Entries marked with a dash (-) indicate that the corresponding uncertainties either do not apply or are less than 1%. The values are given for the 8 TeV analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total error</th>
<th>Stat. error</th>
<th>Expt. syst. error</th>
<th>Theo. syst. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_j = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_{SIG}</td>
<td>16</td>
<td>-</td>
<td>6.7</td>
<td>15</td>
</tr>
<tr>
<td>N_{BKG}</td>
<td>2.5</td>
<td>1.5</td>
<td>1.2</td>
<td>1.7</td>
</tr>
<tr>
<td>N_{WW}</td>
<td>4.2</td>
<td>2.4</td>
<td>2.3</td>
<td>2.6</td>
</tr>
<tr>
<td>N_{TOP}</td>
<td>7.4</td>
<td>2.3</td>
<td>4.2</td>
<td>5.6</td>
</tr>
<tr>
<td>N_{MISID}</td>
<td>17</td>
<td>-</td>
<td>9.9</td>
<td>14</td>
</tr>
<tr>
<td>N_{V}</td>
<td>9.9</td>
<td>4.8</td>
<td>4.6</td>
<td>7.4</td>
</tr>
<tr>
<td>N_{TT} (DY)</td>
<td>34</td>
<td>1.7</td>
<td>33</td>
<td>7.2</td>
</tr>
<tr>
<td>N_{ee/\mu} (DY)</td>
<td>30</td>
<td>14</td>
<td>26</td>
<td>5.5</td>
</tr>
<tr>
<td>n_j = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_{SIG}</td>
<td>22</td>
<td>-</td>
<td>5.3</td>
<td>22</td>
</tr>
<tr>
<td>N_{BKG}</td>
<td>3</td>
<td>1.7</td>
<td>1.4</td>
<td>2.1</td>
</tr>
<tr>
<td>N_{WW}</td>
<td>7.7</td>
<td>5.5</td>
<td>2.7</td>
<td>4.6</td>
</tr>
<tr>
<td>N_{TOP}</td>
<td>5</td>
<td>3.4</td>
<td>2.9</td>
<td>2.3</td>
</tr>
<tr>
<td>N_{MISID}</td>
<td>18</td>
<td>-</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>N_{V}</td>
<td>14</td>
<td>8.9</td>
<td>6.1</td>
<td>8.5</td>
</tr>
<tr>
<td>N_{TT} (DY)</td>
<td>27</td>
<td>3.3</td>
<td>26</td>
<td>6.3</td>
</tr>
<tr>
<td>N_{ee/\mu} (DY)</td>
<td>39</td>
<td>27</td>
<td>26</td>
<td>7.4</td>
</tr>
<tr>
<td>n_j ≥ 2 ggF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_{SIG}</td>
<td>23</td>
<td>-</td>
<td>8.6</td>
<td>22</td>
</tr>
<tr>
<td>N_{BKG}</td>
<td>4.2</td>
<td>1.5</td>
<td>2.2</td>
<td>3.2</td>
</tr>
<tr>
<td>N_{WW}</td>
<td>20</td>
<td>-</td>
<td>8.7</td>
<td>18</td>
</tr>
<tr>
<td>N_{TOP}</td>
<td>7.9</td>
<td>2.6</td>
<td>3.4</td>
<td>6.7</td>
</tr>
<tr>
<td>N_{MISID}</td>
<td>29</td>
<td>-</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>N_{V}</td>
<td>32</td>
<td>-</td>
<td>9.6</td>
<td>31</td>
</tr>
<tr>
<td>N_{TT} (DY)</td>
<td>18</td>
<td>8</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>N_{ee/\mu} (DY)</td>
<td>15</td>
<td>-</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>n_j ≥ 2 VBF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_{SIG}</td>
<td>13</td>
<td>-</td>
<td>6.8</td>
<td>12</td>
</tr>
<tr>
<td>N_{BKG}</td>
<td>9.2</td>
<td>4.7</td>
<td>6.4</td>
<td>4.5</td>
</tr>
<tr>
<td>N_{WW}</td>
<td>32</td>
<td>-</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>N_{TOP}</td>
<td>15</td>
<td>9.6</td>
<td>7.6</td>
<td>8.5</td>
</tr>
<tr>
<td>N_{MISID}</td>
<td>22</td>
<td>-</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>N_{V}</td>
<td>20</td>
<td>-</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>N_{TT} (DY)</td>
<td>40</td>
<td>25</td>
<td>31</td>
<td>2.9</td>
</tr>
<tr>
<td>N_{ee/\mu} (DY)</td>
<td>19</td>
<td>11</td>
<td>15</td>
<td>-</td>
</tr>
</tbody>
</table>

Background contamination in the control regions causes anti-correlations between different background processes, resulting in an uncertainty on the total background smaller than the sum in quadrature of the individual process uncertainties. This effect is called “cross talk” and is most prominent between the WW and top-quark backgrounds in the n_j = 1 category. The uncertainties on the background estimates, as described in Sec. VI, cannot be directly compared to the ones presented in Table XXIII. The latter uncertainties are postfit and are subject to subtle effects, such as the cross talk mentioned above, and also pulls and data constraints (defined below) on the various nuisance parameters.

D. Checks of fit results

The fit simultaneously extracts the signal strength \( \mu \) and the set of auxiliary parameters \( \theta \). This process adjusts the initial prefit estimation of every parameter \( \theta \) as well as its uncertainty, \( \Delta \theta \). The fit model is designed to avoid any significant constraints on the input uncertainties to minimize the assumptions on the correlations between the phase spaces in which they are measured and applied. This is achieved by having mostly single-bin control regions. Of central importance is the prefit and postfit comparison of how the variation of a given systematic source translates to an uncertainty on \( \mu \).

The impact of a single nuisance parameter \( \theta \) is assessed by considering its effect on the signal strength, i.e.,

\[
\Delta \mu_{\pm} = \hat{\mu}(\hat{\theta} \pm \Delta \theta) - \hat{\mu}(\hat{\theta}),
\]

where \( \hat{\mu} \) is the postfit value of the signal strength. In the following, quantities with a hat represent postfit parameter values or their uncertainties.

The values \( \hat{\mu}(\hat{\theta} \pm \Delta \theta) \) are the result of a fit with one \( \theta \) varied by \( \pm \Delta \theta \) around the postfit value for \( \theta \), namely \( \hat{\theta} \). All other \( \theta \) are floating in these fits. In the prefit scenario, the \( \Delta \theta \) are taken as their prefit values of \( \pm 1 \), as \( \theta \) is constrained by a unit Gaussian. The postfit scenario is similar, but with \( \hat{\theta} \) varied by its postfit uncertainty of \( \Delta \hat{\theta} \). This uncertainty is found by a scan about the maximum so that the likelihood ratio takes the values

\[
-2 \ln(\mathcal{L}(\hat{\theta} \pm \Delta \hat{\theta})/\mathcal{L}(\hat{\theta})) = 1.
\]

The corresponding impact on \( \mu \) is \( \Delta \mu_{\hat{\mu}} \).

When \( \Delta \theta \) is less than the prefit value, \( \theta \) is said to be data constrained. In this case the systematic uncertainty is reduced below its input value given the information from the data. This can result from the additional information that the data part of the likelihood injects. As can be seen from Table XXIV, only a few of the uncertainties are data constrained, and only one of them is data constrained by more than 20%. That is the WW generator modeling that includes the m_T shape uncertainties correlated with the uncertainties on the extrapolation factor \( \alpha_{WW} \). The data constraint in this case comes from the high-m_T tail of the

Table XV, because they are largely uncorrelated between lepton p_T bins (statistical uncertainties on the Z + jets data sample) and between the lepton flavors (systematic uncertainties on the OC correction factor). The uncertainty due to the limited sample of background MC events for all the considered processes is included in the experimental component.
signal region, which contains a large fraction of $WW$ events.

The postfit values for $\theta$ modify the rates of signal and background processes, and the data constraints affect the corresponding uncertainties. The results of these shifts are summarized in Table XXIV for a set of $20$ nuisance parameters ordered by the magnitude of $\Delta \mu$ (Higgs signal hypothesis is taken at $m_H = 125$ GeV). The highest-ranked nuisance parameter is the uncertainty on the total ggF cross section due to the PDF variations. It changes $\hat{\mu}$ by $-0.06/0.06$ when varied up and down by $\Delta \theta$, respectively. It is followed by the uncertainty on the total ggF cross section due to QCD scale variations and WW generator modeling uncertainty. Other uncertainties that have a significant impact on $\hat{\mu}$ include the effects of generator modeling on $a_{\text{top}}$, the systematic uncertainties on $a_{\text{miss}}$ originating from a correction for oppositely charged electrons and muons, the luminosity determination for 8 TeV data, and various theoretical uncertainties on the ggF and VBF signal production processes. In total there are 253 nuisance parameters which are divided into three main categories: experimental uncertainties (137 parameters), theoretical uncertainties (72 parameters), and normalization uncertainties (44 parameters). They are further divided into more categories as shown in Table XXVI.

### VIII. YIELDS AND DISTRIBUTIONS

The previous section described the different parameters of the simultaneous fit to the various signal categories defined in the preceding sections. In particular, the signal and background rates and shapes are allowed to vary in order to fit the data in both the signal and control regions, within their associated uncertainties.

In the figures and tables presented in this section, background processes are individually normalized to their postfit rates, which account for changes in the normalization factors ($\beta$) and for pulls of the nuisance parameters ($\theta$). The varying background composition as a function of $m_T$ (or $O_{\text{BVT}}$ in the $n_j \geq 2$ VBF-enriched category) induces a shape uncertainty on the total estimated background. As described in Sec. VII C, additional specific shape uncertainties are included in the fit procedure and are accounted for in the results presented in Sec. IX. No specific $m_T$ shape uncertainties are applied to the figures since their contribution to the total systematic uncertainty band was found to be negligible. The Higgs boson signal rate is normalized to the observed signal strength reported in Sec. IX.

This section is organized as follows. The event yields are presented in Sec. VIII A for each signal category including the statistical and systematic uncertainties. The relevant distributions in the various signal regions are shown in.

### TABLE XXIV. Impact on the signal strength $\hat{\mu}$ from the prefit and postfit variations of the nuisance parameter uncertainties, $\Delta \theta$. The $+(-)$ column header indicates the positive (negative) variation of $\Delta \theta$ and the resulting change in $\hat{\mu}$ is noted in the entry (the sign represents the direction of the change). The right-hand side shows the pull of $\theta$ and the data constraint of $\Delta \theta$. The pulls are given in units of standard deviations ($\sigma$) and $\Delta \theta$ of $\pm 1$ means no data constraint. The rows are ordered by the size of a change in $\hat{\mu}$ due to varying $\theta$ by the postfit uncertainty $\Delta \theta$.

<table>
<thead>
<tr>
<th>Systematic source</th>
<th>Impact on $\hat{\mu}$</th>
<th>Impact on $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggF H, PDF variations on cross section</td>
<td>$-0.06/0.06$</td>
<td>$-0.06 \pm 1$</td>
</tr>
<tr>
<td>ggF H, QCD scale on total cross section</td>
<td>$-0.05/0.05$</td>
<td>$-0.05 \pm 1$</td>
</tr>
<tr>
<td>WW, generator modeling</td>
<td>$-0.07/0.05$</td>
<td>$0 \pm 0.7$</td>
</tr>
<tr>
<td>Top quarks, generator modeling on $a_{\text{top}}$ in ggF cat.</td>
<td>$+0.03/-0.03$</td>
<td>$-0.40 \pm 0.9$</td>
</tr>
<tr>
<td>Misid of $\mu$, OC uncorrelated corr factor $a_{\text{miss}}$, 2012</td>
<td>$-0.03/0.03$</td>
<td>$0.48 \pm 0.8$</td>
</tr>
<tr>
<td>Integrated luminosity, 2012</td>
<td>$-0.03/0.03$</td>
<td>$0.08 \pm 1$</td>
</tr>
<tr>
<td>Misid of $e$, OC uncorrelated corr factor $a_{\text{miss}}$, 2012</td>
<td>$-0.03/0.03$</td>
<td>$-0.06 \pm 0.9$</td>
</tr>
<tr>
<td>ggF H, PDF variations on acceptance</td>
<td>$-0.02/0.02$</td>
<td>$-0.03 \pm 1$</td>
</tr>
<tr>
<td>Jet energy scale, $\eta$ intercalibration</td>
<td>$-0.02/0.02$</td>
<td>$0.45/0.95$</td>
</tr>
<tr>
<td>VBF H, UE/PS</td>
<td>$-0.02/0.02$</td>
<td>$0.26 \pm 1$</td>
</tr>
<tr>
<td>ggF H, QCD scale on $e_l$</td>
<td>$-0.01/0.01$</td>
<td>$-0.10 \pm 0.95$</td>
</tr>
<tr>
<td>Muon isolation efficiency</td>
<td>$-0.02/0.02$</td>
<td>$0.13 \pm 1$</td>
</tr>
<tr>
<td>VBF QCD scale on acceptance</td>
<td>$-0.02/0.02$</td>
<td>$0.09 \pm 1$</td>
</tr>
<tr>
<td>ggF H, UE/PS</td>
<td>$-/-0.02$</td>
<td>$0 \pm 0.9$</td>
</tr>
<tr>
<td>ggF H, QCD scale on acceptance</td>
<td>$-0.02/0.02$</td>
<td>$0 \pm 1$</td>
</tr>
<tr>
<td>Light jets, tagging efficiency</td>
<td>$+0.02/0.02$</td>
<td>$0.21 \pm 1$</td>
</tr>
<tr>
<td>ggF H, generator modeling on acceptance</td>
<td>$+0.01/-0.01$</td>
<td>$0.10 \pm 1$</td>
</tr>
<tr>
<td>ggF H, QCD scale on $n_j \geq 2$ cross section</td>
<td>$-0.01/0.01$</td>
<td>$-0.04 \pm 1$</td>
</tr>
<tr>
<td>Top quarks, generator modeling on $a_{\text{top}}$ in VBF cat.</td>
<td>$-0.01/0.01$</td>
<td>$-0.16 \pm 1$</td>
</tr>
<tr>
<td>Electron isolation efficiency</td>
<td>$-0.02/0.02$</td>
<td>$-0.14 \pm 1$</td>
</tr>
</tbody>
</table>

-0.1 -0.05 0 0.05 0.1
TABLE XXV. Signal region yields with uncertainties. The tables give the ggF- and VBF-enriched postfit yields for each $n_j$ category, separated for the 8 and 7 TeV data analyses. The $N_{signal}$ columns show the expected signal yields from the ggF and VBF production modes, with values scaled to the observed combined signal strength (see Sec. IX C). For each group separated of rows separated by a vertical gap, the first line gives the combined values for the different subchannels or BDT bins. The yields and the uncertainties take into account the pulls and data constraints of the nuisance parameters, and the correlations between the channels and the background categories. The quoted uncertainties include the theoretical and experimental systematic sources and those due to sample statistics. Values less than 0.1 (0.01) events are written as 0.0 (-).

<table>
<thead>
<tr>
<th>Channel</th>
<th>$N_{obs}$</th>
<th>$N_{bkg}$</th>
<th>$N_{signal}$</th>
<th>$N_WW$</th>
<th>$N_t$</th>
<th>$N_{Wj}$</th>
<th>$N_{Wj}^{misid}$</th>
<th>$N_{jj}$</th>
<th>$N_{VV}$</th>
<th>$N_{DY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_j = 0$</td>
<td>3750</td>
<td>3430 ± 90</td>
<td>300 ± 500</td>
<td>8 ± 4</td>
<td>2250 ± 95</td>
<td>112 ± 9</td>
<td>195 ± 15</td>
<td>360 ± 60</td>
<td>16 ± 5</td>
<td>420 ± 40</td>
</tr>
<tr>
<td>$\mu\mu$, $\ell_2 = \mu$</td>
<td>1430</td>
<td>1280 ± 40</td>
<td>129 ± 20</td>
<td>3.0 ± 2.1</td>
<td>830 ± 34</td>
<td>41 ± 3</td>
<td>73 ± 6</td>
<td>149 ± 29</td>
<td>10.1 ± 3.6</td>
<td>167 ± 21</td>
</tr>
<tr>
<td>$ee$, $\ell_2 = e$</td>
<td>1212</td>
<td>1106 ± 35</td>
<td>97 ± 15</td>
<td>2.5 ± 0.6</td>
<td>686 ± 29</td>
<td>33 ± 3</td>
<td>57 ± 5</td>
<td>128 ± 31</td>
<td>3.8 ± 1.5</td>
<td>184 ± 23</td>
</tr>
<tr>
<td>$ee/\mu\mu$</td>
<td>1108</td>
<td>1040 ± 40</td>
<td>77 ± 15</td>
<td>2.4 ± 1.7</td>
<td>740 ± 40</td>
<td>39 ± 3</td>
<td>65 ± 5</td>
<td>82 ± 16</td>
<td>2 ± 0.5</td>
<td>68 ± 7</td>
</tr>
<tr>
<td>$n_j = 1$</td>
<td>1596</td>
<td>1470 ± 40</td>
<td>102 ± 26</td>
<td>17 ± 5</td>
<td>630 ± 50</td>
<td>150 ± 10</td>
<td>385 ± 20</td>
<td>108 ± 20</td>
<td>8.2 ± 3.0</td>
<td>143 ± 20</td>
</tr>
<tr>
<td>$\mu\mu$, $\ell_2 = \mu$</td>
<td>621</td>
<td>569 ± 19</td>
<td>45 ± 11</td>
<td>7.4 ± 2</td>
<td>241 ± 20</td>
<td>58 ± 4</td>
<td>147 ± 7</td>
<td>51 ± 11</td>
<td>5.7 ± 2.0</td>
<td>53 ± 10</td>
</tr>
<tr>
<td>$ee$, $\ell_2 = e$</td>
<td>508</td>
<td>475 ± 18</td>
<td>35 ± 9</td>
<td>6.1 ± 1.4</td>
<td>202 ± 17</td>
<td>45 ± 3</td>
<td>119 ± 6</td>
<td>37 ± 9</td>
<td>2.3 ± 0.9</td>
<td>60 ± 10</td>
</tr>
<tr>
<td>$ee/\mu\mu$</td>
<td>467</td>
<td>427 ± 21</td>
<td>22 ± 6</td>
<td>3.6 ± 1.8</td>
<td>184 ± 15</td>
<td>46 ± 4</td>
<td>119 ± 10</td>
<td>19 ± 4</td>
<td>0.2 ± 0.1</td>
<td>31 ± 4</td>
</tr>
<tr>
<td>$n_j \geq 2$, ggF $\mu\mu$</td>
<td>1017</td>
<td>960 ± 40</td>
<td>37 ± 11</td>
<td>13 ± 1.4</td>
<td>138 ± 28</td>
<td>56 ± 5</td>
<td>480 ± 40</td>
<td>54 ± 25</td>
<td>62 ± 22</td>
<td>56 ± 18</td>
</tr>
<tr>
<td>$n_j \geq 2$, VBF $\mu\mu$</td>
<td>130</td>
<td>99 ± 9</td>
<td>7.7 ± 2.6</td>
<td>21 ± 3</td>
<td>11 ± 3.5</td>
<td>5.5 ± 0.7</td>
<td>29 ± 5</td>
<td>4.7 ± 1.4</td>
<td>2.8 ± 1.0</td>
<td>4.4 ± 0.9</td>
</tr>
<tr>
<td>$ee$, $\mu$ bin 1</td>
<td>37</td>
<td>36 ± 4</td>
<td>3.3 ± 1.2</td>
<td>4.9 ± 0.5</td>
<td>5.0 ± 1.5</td>
<td>3.0 ± 0.6</td>
<td>15.6 ± 2.6</td>
<td>3.2 ± 1.0</td>
<td>2.3 ± 0.8</td>
<td>2.3 ± 0.7</td>
</tr>
<tr>
<td>$ee$, $\mu$ bin 2</td>
<td>14</td>
<td>6.5 ± 1.3</td>
<td>1.4 ± 0.5</td>
<td>4.9 ± 0.5</td>
<td>1.7 ± 0.7</td>
<td>0.3 ± 0.4</td>
<td>2.0 ± 1.0</td>
<td>0.4 ± 0.1</td>
<td>0.3 ± 0.1</td>
<td>0.7 ± 0.2</td>
</tr>
<tr>
<td>$ee$, $\mu$ bin 3</td>
<td>6</td>
<td>1.2 ± 0.3</td>
<td>0.4 ± 0.3</td>
<td>3.8 ± 0.7</td>
<td>0.3 ± 0.1</td>
<td>0.1 ± 0.0</td>
<td>0.3 ± 0.1</td>
<td>-</td>
<td>-</td>
<td>0.1 ± 0.0</td>
</tr>
<tr>
<td>$ee/\mu\mu$, $\mu$ bin 1</td>
<td>53</td>
<td>46 ± 6</td>
<td>1.7 ± 0.6</td>
<td>2.6 ± 0.3</td>
<td>3.1 ± 1.0</td>
<td>1.7 ± 0.3</td>
<td>10.1 ± 1.6</td>
<td>0.9 ± 0.2</td>
<td>0.2 ± 0.1</td>
<td>1.0 ± 0.3</td>
</tr>
<tr>
<td>$ee/\mu\mu$ bin 2</td>
<td>14</td>
<td>8.4 ± 1.8</td>
<td>0.7 ± 0.3</td>
<td>3.0 ± 0.4</td>
<td>0.9 ± 0.3</td>
<td>0.3 ± 0.2</td>
<td>1.2 ± 0.5</td>
<td>0.2 ± 0.1</td>
<td>-</td>
<td>0.3 ± 0.1</td>
</tr>
<tr>
<td>$ee/\mu\mu$ bin 3</td>
<td>6</td>
<td>1.1 ± 0.4</td>
<td>0.2 ± 0.2</td>
<td>2.1 ± 0.4</td>
<td>0.1 ± 0.1</td>
<td>0.1 ± 0.0</td>
<td>0.2 ± 0.1</td>
<td>0.1 ± 0.0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Sec. VIII B. Section VIII C summarizes the differences in the event and object selection, the signal treatment and the background estimates with respect to the previously published analysis [5].

A. Event yields

Table XXV shows the postfit yields for all of the fitted categories in the 8 TeV [Table XXV(a)] and 7 TeV [Table XXV(b)] data analyses. The signal yields are scaled with the observed signal strength derived from the simultaneous combined fit to all of the categories. All of the background processes are normalized to the postfit $\beta$ values (where applicable) and additionally their rates take into account the pulls of the nuisance parameters. The observed and expected yields are shown, for each $n_j$ category, separately for the $ee$ and $ee/\mu\mu$ channels. The sum of the expected and observed yields is also reported. The uncertainties include both the statistical and systematic components.
As described in the previous section, the changes in the normalization factors and the pulls of the nuisance parameters can affect the expected rates of the signal and background processes. The differences between the prefit (tables in Sec. IV) and postfit (Table XXV) expected rates for each background process are compared to the total uncertainty on that expected background, yielding a significance of the change. In the analysis of the \( n_j \leq 1 \) category of the 8 TeV data, most of the changes are well below one standard deviation. In the \( e\mu \) \( n_j = 0 \) sample, the expected multijet background is increased by 1.3 standard deviations (equivalent to a 30% increase in the expected multijet background prediction which corresponds to 2% of the signal prediction) due to the positive pulls of the three nuisance parameters assigned to the uncertainties on the extrapolation factor. A negative pull of the nuisance parameter associated with the uncertainties on the DY \( f_\text{recoll} \) selection efficiency changes the \( Z/\gamma \to ee, \mu\mu \) yield in the \( ee/\mu\mu \) \( n_j = 0 \) sample by 1.6 standard deviations (equivalent to a 40% decrease in DY in this category which corresponds to 25% of the signal prediction).

**B. Distributions**

The transverse mass formed from the dilepton and missing transverse momenta (\( m_T \)) is used as the final discriminant in the extraction of the signal strength in the \( n_j \leq 1 \) and \( n_j \geq 2 \) ggF-enriched categories. The likelihood fit exploits the differences in \( m_T \) shapes between the signal and background processes.

Several of the \( m_T \) distributions for the \( e\mu \) sample (corresponding to different choices of the \( m_{eff} \) and \( p_T^2 \) bins) in the \( n_j \leq 1 \) categories are shown in Fig. 29. The background composition, signal contribution, and the separation in the \( m_T \) distributions between signal and background are different for each region. In general, as shown in Figs. 29(a)-(c), the \( WW \) process dominates the background contributions in regions with \( n_j = 0 \); the difference between these distributions is due to the varying signal contribution and background \( m_T \) shape. In contrast, Fig. 29(d) shows that \( VV \) and \( W + \) jets processes are dominant backgrounds in the \( 10 < m_{eff} < 30 \) GeV and \( 10 < p_T^2 < 15 \) GeV region. For most of the distributions shown in Fig. 29, agreement between data and MC is improved qualitatively when including the expected signal from a Standard Model Higgs boson with \( m_H = 125 \) GeV.

The \( m_T \) distributions for the \( ee/\mu\mu \) samples in the \( n_j \leq 1 \) categories are shown in Fig. 30. In contrast to the \( e\mu \) distributions, the residual DY background is present in these samples at low values of \( m_T \).

For the ggF-enriched \( n_j \geq 2 \) category, Fig. 31 shows the \( m_T \) distribution. In contrast to the \( n_j \leq 1 \) distributions, the dominant backgrounds arise from top-quark and \( Z/\gamma \to \tau\tau \) production (shown together with the negligible contribution from \( Z/\gamma \to ee, \mu\mu \)).

---

**FIG. 29 (color online).** Postfit transverse mass distributions in the \( e\mu \) \( n_j \leq 1 \) categories in the 8 TeV data analysis, for specific \( m_{eff} \) and \( p_T^2 \) ranges. The plots are made after applying all the selection requirements (see Tables V and VI). The signal processes are scaled with the observed signal strength \( \rho \) from the fit to all the regions and the background normalizations include the postfit \( \beta \) values and effects from the pulls of the nuisance parameters. See Fig. 5 for plotting details.
For the VBF-enriched $n_j \geq 2$ category, a selection-based analysis, which uses the $m_T$ distribution as the discriminant, is used as a cross-check of the BDT result. In this case, $m_T$ is divided into three bins (with boundaries at 80 and 130 GeV) and an additional division in $m_{jj}$ at 1 TeV is used in the $e\mu$ channel to profit from the difference in shapes between signal and background processes. Figure 32(a) shows the $m_T$ distribution before the division into the high- and low-$m_{jj}$ regions. Figure 32(b) shows the scatter plot of $m_{jj}$ versus $m_T$. The areas with the highest signal-to-background ratio are characterized by low $m_T$ and high $m_{jj}$.

Figures 33(a) and 33(c) show the $O_{\text{BDT}}$ outputs in the $e\mu$ and $ee/\mu\mu$ samples, respectively. In terms of VBF signal production, the third BDT bin provides the highest purity, with a signal-to-background ratio of approximately 2. The $m_T$ variable is an input to the BDT and its distributions after the BDT classification are shown in Figs. 33(b) and 33(d), combining all three BDT bins, for the $e\mu$ and $ee/\mu\mu$ samples, respectively.

Figure 34 shows the $m_T$ distributions in the 7 TeV analysis in the various signal regions in the $n_j \leq 1$ categories. Characteristics similar to those in the 8 TeV analysis are observed, but with fewer events.

FIG. 30 (color online). Postfit transverse mass distributions in the $n_j \leq 1$, $ee/\mu\mu$ categories in the 8 TeV analysis. See Figs. 5 and 29 for plotting details.

FIG. 31 (color online). Postfit transverse mass distribution in the $n_j \geq 2$ ggF-enriched category in the 8 TeV analysis. See Figs. 5 and 29 for plotting details.

FIG. 32 (color online). Postfit distributions in the cross-check analysis in the $n_j \geq 2$ VBF-enriched category in the 8 TeV data analysis: (a) $m_T$ and (b) $m_{jj}$ versus $m_T$ scatter plot for data. For each bin in (b), the ratio $N_{\text{VBF}}/N_{\text{rest}}$ is stated in the plot, where $N_{\text{rest}}$ includes all processes other than the VBF signal. See Figs. 5 and 29 for plotting details.
Finally, Fig. 35(a) shows the combined $m_T$ distribution, summed over the lepton-flavor samples and the $n_j \leq 1$ categories for the 7 and 8 TeV data analyses. To illustrate the significance of the excess of events observed in data with respect to the total background, the systematic uncertainty on the signal is omitted. The uncertainty band accounts for the correlations between the signal regions, including between the 7 and 8 TeV data, and for the varying size of the uncertainties as a function of $m_T$. Figure 35(b) shows the residuals of the data with respect to the total estimated background compared to the expected $m_T$ distribution of a SM Higgs boson with $m_H = 125$ GeV scaled by the observed combined signal strength (see Sec. IX). The level of agreement observed in Fig. 35(b) between the background-subtracted data and the expected Higgs boson signal strengthens the interpretation of the observed excess as a signal from Higgs boson decay.

C. Differences with respect to previous results

The analysis presented in this paper has better sensitivity than the previous ATLAS analysis [5]. The most important changes—described in detail below—include improvements in the object identification, the signal acceptance, the background estimation and modeling, and the fit procedure.

FIG. 33 (color online). Postfit BDT and transverse mass distributions in the $n_j \geq 2$ VBF-enriched category in the 8 TeV data analysis: (a) BDT output in $e\mu$, (b) $m_T$ in $e\mu$, (c) BDT output in $ee/\mu\mu$, and (d) $m_T$ in $ee/\mu\mu$. For (b) and (d), the three BDT bins are combined. See Figs. 5 and 29 for plotting details.

FIG. 34 (color online). Postfit transverse mass distributions in the $n_j \leq 1$ categories in the 7 TeV data analysis, for specific $m_{ee}$ and $p_T^2$ ranges. The plots are made after applying all the selection requirements (see Sec. IV E). See Figs. 5 and 29 for plotting details.
The methods used to estimate nearly all of the background contributions in the signal region are improved. These improvements lead to a better understanding of the normalizations and thus the systematic uncertainties. The rejection of the top-quark background is improved by applying a veto on $b$-jets with $p_T > 20$ GeV, which is below the nominal 25 GeV threshold in the analysis. A new method of estimating the jet $b$-tagging efficiency is used. It results in the cancellation of the $b$-tagging uncertainties between the top-quark control region and signal regions in the $n_j = 1$ categories. The $Z/\gamma^* \rightarrow \tau\tau$ background process is normalized to the data in a dedicated high-statistics control region in the $n_j \leq 1$ and $n_j \geq 2$ ggF-enriched categories. The VV backgrounds are normalized to the data using a new control region, based on a sample with two same-charge leptons. Introducing this new control region results in the cancellation of most of the theoretical uncertainties on the VV backgrounds. The multijet background is now explicitly estimated with an extrapolation factor method using a sample with two anti-identified leptons. Its contribution is negligible in the $n_j \leq 1$ category, but it is at the same level as $W +$ jets background in the $n_j \geq 2$ ggF-enriched category. A large number of improvements are applied to the estimation of the $W +$ jets background, one of them being an estimation of the extrapolation factor using $Z +$ jets instead of dijet data events.

Signal yield uncertainties are smaller than in the previous analysis. The uncertainties on the jet multiplicity distribution in the ggF signal sample, previously estimated with the Stewart-Tackmann technique [80], are now estimated with the jet-veto-efficiency method [79]. This method yields more precise estimates of the signal rates in the exclusive jet bins in which the analysis is performed.

The $n_j \geq 2$ sample is divided into VBF- and ggF-enriched categories. The BDT technique, rather than a selection-based approach, is used for the VBF category. This improves the sensitivity of the expected VBF results by 60% relative to the previously published analysis. The ggF-enriched category is a new subcategory that targets ggF signal production in this sample.

In summary, the analysis presented in this paper brings a gain of 50% in the expected significance relative to the previous published analysis [5].

**IX. RESULTS AND INTERPRETATIONS**

Combining the 2011 and 2012 data in all categories, a clear excess of signal over the background is seen in Fig. 35. The profile likelihood fit described in Sec. VII B is used to search for a signal and characterize the production rate in the ggF and VBF modes. Observation of the inclusive Higgs boson signal, and evidence for the VBF production mode, are established first. Following that, the excess in data is characterized using the SM Higgs boson as the signal hypothesis, up to linear rescalings of the production cross sections and decay modes. Results include...
the inclusive signal strength as well as those for the individual ggF and VBF modes. This information is also interpreted as a measurement of the vector-boson and fermion couplings of the Higgs boson, under the assumptions outlined in Ref. [67]. Because this is the first observation in the $WW^* \rightarrow \ell\nu\ell\nu$ channel using ATLAS data, the exclusion sensitivity and observed exclusion limits as a function of $m_H$ are also presented to illustrate the improvements with respect to the version of this analysis used in the 2012 discovery [4]. Finally, cross-section measurements, both inclusive and in specific fiducial regions, are presented. All results in this section are quoted for a Higgs boson mass corresponding to the central value of the ATLAS measurement in the $ZZ \rightarrow 4\ell$ and $\gamma\gamma$ decay modes, $m_H = 125.36 \pm 0.41$ GeV [9].

### A. Observation of the $H \rightarrow WW^*$ decay mode

The test statistic $q_\mu$, defined in Sec. VII B, is used to quantify the significance of the excess observed in Sec. VIII. The probability that the background can fluctuate to produce an excess at least as large as the one observed in data is called $p_0$ and is computed using $q_\mu$ with $\mu = 0$. It depends on the mass hypothesis $m_H$ through the distribution used to extract the signal ($m_T$ or $O_{BR_T}$). The observed and expected $p_0$ are shown as a function of $m_H$ in Fig. 36. The observed curve presents a broad minimum centered around $m_H \approx 130$ GeV, in contrast with the higher $p_0$ values observed for lower and higher values of $m_H$. The shapes of the observed and expected curves are in good agreement.

The probability $p_0$ can equivalently be expressed in terms of the number of standard deviations, referred to as the local significance ($Z_0$ defined in Sec. VII B 2). The value of $p_0$ as a function of $m_H$ is found by scanning $m_H$ in 5 GeV intervals. The minimum $p_0$ value is found at $m_H = 130$ GeV and corresponds to a local significance of 6.1 standard deviations. The same observed significance within the quoted precision is found for $m_H = 125.36$ GeV. This result establishes a discovery-level signal in the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ channel alone. The expected significance for a SM Higgs boson at the same mass is 5.8 standard deviations.

In order to assess the compatibility with the SM expectation for a Higgs boson of mass $m_H$, the observed best-fit $\mu$ value as a function of $m_H$ is shown in Fig. 37. The observed $\mu$ is close to zero for $m_H > 160$ GeV and crosses unity at $m_H \approx 125$ GeV. The increase of $\mu$ for small values of $m_H$ is expected in the presence of a signal with mass $m_H = 125.36$ GeV, as is also shown in Fig. 37. The dependence of $\mu$ on the value of $m_H$ ($d\mu/dm_H = -0.078$ GeV at $m_H = 125.36$ GeV) arises mostly from the dependence of the $WW^*$ branching fraction on $m_H$.

The assumption that the total yield is predicted by the SM is relaxed to evaluate the two-dimensional likelihood contours of $(m_H, \mu)$, shown in Fig. 38. The value ($\mu = 1, m_H = 125.36$ GeV) lies well within the 68% C.L. contour, showing that the signal observed is compatible with those in the high-resolution channels.

### B. Evidence for VBF production

The $n_\ell \geq 2$ VBF-enriched signal region was optimized for its specific sensitivity to the VBF production process, as described in particular in Sec. IV. Nevertheless, as can be seen in Table XXV, the ggF contribution to this signal region is large, approximately 30%, so it has to be profiled by the global fit together with the extraction of the significance of the signal strength of the VBF production process.

![FIG. 36](color online). Local $p_0$ as a function of $m_H$. The observed values are shown as a solid line with points where $p_0$ is evaluated. The dashed line shows the expected values given the presence of a signal at each x-axis value. The expected values for $m_H = 125.36$ GeV are given as a solid line without points; the inner (outer) band shaded darker (lighter) represents the one (two) standard deviation uncertainty.

![FIG. 37](color online). Best-fit signal strength $\mu$ as a function of $m_H$. The observed values are shown as a solid line with points where $\mu$ is evaluated. The expected values for $m_H = 125.36$ GeV are shown as a solid line without points. The dashed and shaded (solid) bands represent the one standard deviation uncertainties for the observed (expected) values.
The global likelihood can be evaluated as a function of the ratio $\mu_{\text{VBF}}/\mu_{\text{ggF}}$, with both signal strengths varied independently. The result is illustrated in Fig. 39, which has a best-fit value for the ratio of

$$
\frac{\mu_{\text{VBF}}}{\mu_{\text{ggF}}} = 1.26^{+0.61}_{-0.45} \text{(stat)} ^{+0.50}_{-0.26} \text{(syst)} = 1.26^{+0.79}_{-0.53}.
$$

The value of the likelihood at $\mu_{\text{VBF}}/\mu_{\text{ggF}} = 0$ can be interpreted as the observed significance of the VBF production process for $m_H = 125.36$ GeV, and corresponds to 3.2 standard deviations; the expected significance is 2.7 standard deviations. This establishes the evidence for the VBF production mode in the $H \rightarrow WW^{*} \rightarrow \ell\nu\ell\nu$ final state. The significance derived from testing the ratio $\mu_{\text{VBF}}/\mu_{\text{ggF}} = 0$ is equivalent to the significance of testing $\mu_{\text{VBF}} = 0$, though testing the ratio is conceptually advantageous since the branching fraction cancels in this parameter, while it is implicit in $\mu_{\text{VBF}}$.

This result was verified with the cross-check analysis described in Sec. IV C, in which the multivariate discriminant is replaced with a series of event selection requirements motivated by the VBF topology. The expected and observed significances at $m_H = 125.36$ GeV are 2.1 and 3.0 standard deviations, respectively. The compatibility of the 8 TeV results from the cross-check and $O_{\text{BDT}}$ analyses was checked with pseudoexperiments, considering the statistical uncertainties only and fixing $\mu_{\text{ggF}}$ to 1.0. With those caveats, the probability that the difference in $Z_0$ values is larger than the one observed is 79%, reflecting good agreement.

### C. Signal strength $\mu$

The parameter $\mu$ is used to characterize the inclusive Higgs boson signal strength as well as subsets of the signal regions or individual production modes. First, the ggF and VBF processes can be distinguished by using the normalization parameter $\mu_{\text{ggF}}$ for the signal predicted for the ggF signal process, and $\mu_{\text{VBF}}$ for the signal predicted for the VBF signal process. This can be done for a fit to any set of the signal regions in the various categories. In addition, to check that the measured value is consistent among categories, different subsets of the signal regions can be fit. For example, the $n_j = 0$ and $n_j = 1$ categories can be compared, or the $e\mu$ and $ee/\mu\mu$ categories. To derive these results, only the signal regions are separated; the control region definitions do not change. In particular, the control regions defined using only $e\mu$ events are used, even when only $ee/\mu\mu$ signal regions are considered.

The combined Higgs signal strength $\mu$, including 7 and 8 TeV data and all signal region categories, is

$$
\mu = 1.09 \pm 0.16 \text{ (stat)} ^{+0.08}_{-0.07} \text{ (exp)} ^{+0.05}_{-0.12} \text{ (syst)} ^{+0.15}_{-0.12} \text{ (theo)} ^{+0.03}_{-0.01} \text{ (lumi)}
$$

$$
= 1.09 \pm 0.15 \text{ (stat)} ^{+0.17}_{-0.14} \text{ (syst)}
$$

$$
= 1.09 \pm 0.21.
$$

FIG. 39 (color online). Likelihood scan as a function of $\mu_{\text{VBF}}/\mu_{\text{ggF}}$ for $m_H = 125.36$ GeV. The value of the likelihood at $\mu_{\text{VBF}}/\mu_{\text{ggF}} = 0$ gives the significance of the VBF signal at 3.2 standard deviations. The inner (middle) [outer] band shaded darker (lighter) [darker] represents the one (two) [three] standard deviation uncertainty around the central value represented by the vertical line.

FIG. 38 (color online). Observed signal strength $\mu$ as a function of $m_H$ as evaluated by the likelihood fit. The shaded areas represent the one, two, and three standard deviation contours with respect to the best-fit values $m_{\mu}$ and $\mu$.
The statistical uncertainty accounts for the number of observed events in the signal regions and profiled control regions. The statistical uncertainties from Monte Carlo simulated samples, from nonprofiled control regions, and from the extrapolation factors used in the W + jets background estimate are all included in the experimental uncertainties here and for all results in this section. The theoretical uncertainty includes uncertainties on the signal acceptance and cross section as well as theoretical uncertainties on the background extrapolation factors and normalizations. The expected value of $\mu$ is 1 + 0.16(stat) ± 0.17(syst).

In order to check the compatibility with the SM predictions of the ggF and VBF production processes, $\mu_{ggF}$ and $\mu_{VBF}$ can be simultaneously determined through a fit to all categories because of the different sensitivity to these processes in the various categories. In this fit, the VH contribution is included although there is no dedicated category for it, and the SM value for the ratio $\sigma_{VBF}/\sigma_{VH}$ is assumed. Technically, the signal strength $\mu_{VBF+VH}$ is measured, but because the contribution from VH is negligible, the notation $\mu_{VBF}$ is used. The corresponding two-dimensional likelihood contours as a function of $\mu_{ggF}$ and $\mu_{VBF}$ are shown in Fig. 40. Using the same treatment, the separate signal strengths can be measured. The results are

$$\mu_{ggF} = 1.02 \pm 0.19$$

$$\mu_{VBF} = 1.27 \pm 0.44$$

For all of these results, the signal acceptance for all categories is consistent with each other and with unity within the assigned uncertainties. For the overall signal strength, the contribution from the $n_j \geq 2$ VBF category is second only to the $n_j = 0$ ggF category, and the $n_j \geq 2$ ggF category contribution is comparable to those in the $n_j = 0$ and $n_j = 1$ ee/$\mu\mu$ categories.

For all of these results, the signal acceptance for all production modes is evaluated assuming a SM Higgs boson. The VH production process contributes a small number of events, amounting to about 1% of the expected signal from the VBF process. It is included in the predicted signal yield, and where relevant, is grouped with the VBF signal assuming the SM value of the ratio $\sigma_{VBF}/\sigma_{VH}$. The small (<1%) contribution of $H \rightarrow \tau\tau$ to the signal regions is treated as signal, assuming the branching fractions as predicted by the SM.

D. Higgs couplings to fermions and vector bosons

The values of $\mu_{ggF}$ and $\mu_{VBF}$ can be used to test the compatibility of the fermionic and bosonic couplings of the Higgs boson with the SM prediction using a framework motivated by the leading-order interactions [67]. The parametrization uses the scale factors $\kappa_F$, applied to all fermionic couplings, and $\kappa_V$, applied to all bosonic couplings; these parameters are unity for the SM.

In particular, the ggF production cross section is proportional to $\kappa_F^2$ through the top-quark or bottom-quark loops at the production vertex, and the VBF production cross section is proportional to $\kappa_V^4$. The branching fraction $B_{H \rightarrow WW}$ is proportional to $\kappa_F^2$ and inversely proportional to a linear combination of $\kappa_F^2$ and $\kappa_V^4$. This model assumes that there are no non-SM decay modes, so the denominator corresponds to the total decay width in terms of the fermionic and bosonic decay amplitudes. The formulas, following Ref. [67], are

$$\mu_{ggF} \propto \frac{\kappa_F^2 \cdot \kappa_V^4}{(B_{H \rightarrow f f} + B_{H \rightarrow W W})\kappa_F^2 + (B_{H \rightarrow V V} \kappa_V^4)}$$

$$\mu_{VBF} \propto \frac{\kappa_V^4}{(B_{H \rightarrow f f} + B_{H \rightarrow W W})\kappa_F^2 + (B_{H \rightarrow V V} \kappa_V^4)}$$

The small contribution from $B_{H \rightarrow \tau\tau}$ depends on both $\kappa_F$ and $\kappa_V$ and is not explicitly shown. Because $(B_{H \rightarrow f f} + B_{H \rightarrow W W}) \approx 0.75$, $\kappa_F^2$ is the dominant component of the denominator.
for $\kappa_F^2 \gtrsim \frac{1}{2} \kappa_V^2$. As a result, the $\kappa_F^2$ dependence for the ggF process approximately cancels, but the rate remains sensitive to $\kappa_V$. Similarly, the VBF rate scales approximately with $\kappa_V^4/\kappa_F^4$ and the VBF channel provides more sensitivity to $\kappa_F$ than the ggF channel does in this model. Because Eq. (17) contains only $\kappa_F^2$ and $\kappa_V^2$, this channel is not sensitive to the sign of $\kappa_F$ or $\kappa_V$.

The likelihood scan as a function of $\kappa_V$ and $\kappa_F$ is shown in Fig. 41. Both the observed and expected contours are shown, and are in good agreement. The relatively low discrimination among high values of $\kappa_F$ in the plot is due to the functional behavior of the total ggF yield. The product $\sigma_{ggF} \cdot B$ does not depend on $\kappa_F$ in the limit where $\kappa_F \gg \kappa_V$, so the sensitivity at high $\kappa_F$ values is driven by the value of $\mu_{VBF}$. The VBF process rapidly vanishes in the limit where $\kappa_F \gg \kappa_V$ due to the increase of the Higgs boson total width and the consequent reduction of the branching fraction to WW bosons. Therefore, within this framework, excluding $\mu_{VBF} = 0$ excludes $\kappa_F \gg \kappa_V$.

The best fit values are

$$\kappa_F = 0.93 \pm 0.24 \pm 0.21 = 0.93 \pm 0.32 \pm 0.23$$

$$\kappa_V = 1.04 \pm 0.07 \pm 0.07 = 1.04 \pm 0.11$$

and their correlation is $\rho = 0.47$. The correlation is derived from the covariance matrix constructed from the second-order mixed partial derivatives of the likelihood, evaluated at the best-fit values of $\kappa_F$ and $\kappa_V$.

---

**TABLE XXVI. Summary of uncertainties on the signal strength $\mu$.** The table gives the relative uncertainties for inclusive Higgs production (left), ggF production (middle), and VBF production (right). For each group separated by a horizontal line, the first line gives the combined result. The “profiled signal region” indicates the contribution of the uncertainty on the ggF signal yield to the $\mu_{VBF}$ measurement and vice versa. The “misid” is the systematic uncertainty related to the $W + jets$ background estimation. The “$Z/\gamma^* \rightarrow ee, \mu\mu$” entry corresponds to uncertainties on the $f_{\text{recol}}$ selection efficiency for the $n_j \leq 1 ee/\mu\mu$ category. The “muons and electrons” entry includes uncertainties on the lepton energy scale, lepton momentum corrections, lepton trigger efficiencies, and lepton isolation efficiencies. The “jets” uncertainties include the jet energy scale, jet energy resolution, and the $b$-tagging efficiency. Values are quoted assuming $m_H = 125.36$ GeV. The plot for VBF (third column) has a different scale than the other columns to show the relative uncertainties per column. The entries marked with a dash are smaller than 0.01 or do not apply.

<table>
<thead>
<tr>
<th>Source</th>
<th>Observed $\mu = 1.09$</th>
<th>Observed $\mu_{ggF} = 1.02$</th>
<th>Observed $\mu_{VBF} = 1.27$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error (scaled by 100)</td>
<td>Plot of error</td>
<td>Error (scaled by 100)</td>
</tr>
<tr>
<td>Data statistics</td>
<td>0.16 0.15</td>
<td>-</td>
<td>0.19 0.19</td>
</tr>
<tr>
<td>Signal regions</td>
<td>0.12 0.12</td>
<td>-</td>
<td>0.14 0.14</td>
</tr>
<tr>
<td>Profiled control regions</td>
<td>0.10 0.10</td>
<td>-</td>
<td>0.12 0.12</td>
</tr>
<tr>
<td>Profiled signal regions</td>
<td>0.03 0.03</td>
<td>-</td>
<td>0.09 0.08</td>
</tr>
<tr>
<td>MC statistics</td>
<td>0.04 0.04</td>
<td>-</td>
<td>0.06 0.06</td>
</tr>
<tr>
<td>Theoretical systematics</td>
<td>0.15 0.12</td>
<td>-</td>
<td>0.19 0.16</td>
</tr>
<tr>
<td>Background WW</td>
<td>0.06 0.06</td>
<td>-</td>
<td>0.08 0.08</td>
</tr>
<tr>
<td>Background top quark</td>
<td>0.03 0.03</td>
<td>-</td>
<td>0.04 0.04</td>
</tr>
<tr>
<td>Background misid. factor</td>
<td>0.05 0.05</td>
<td>-</td>
<td>0.06 0.06</td>
</tr>
<tr>
<td>Others</td>
<td>0.02 0.02</td>
<td>-</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>0.03 0.03</td>
<td>-</td>
<td>0.03 0.02</td>
</tr>
<tr>
<td>Total</td>
<td>0.23 0.21</td>
<td>-</td>
<td>0.29 0.26</td>
</tr>
</tbody>
</table>

-30 -15 0 15 30

**G. Aad et al.**

*PHYSICAL REVIEW D 92, 012006 (2015)*
E. Exclusion limits

The analysis presented in this paper has been optimized for a Higgs boson of mass \( m_H = 125 \) GeV, but, due to the low mass resolution of the \( \ell \nu \ell \nu \) channel, it is sensitive to SM-like Higgs bosons of mass up to 200 GeV and above. The exclusion ranges are computed using the modified frequentist method \( \text{CL}_S [99] \). A SM Higgs boson of mass \( m_H = 125.36 \) GeV is considered excluded at 95\% C.L. if the value \( \mu = 1 \) is excluded at that mass. The analysis is expected to exclude a SM Higgs boson with mass down to 114 GeV at 95\% C.L. The clear excess of signal over background, shown in the previous sections, results in an observed exclusion range of \( 132 < m_H < 200 \) GeV, extending to the upper limit of the search range, as shown in Fig. 42.

F. Higgs boson production cross sections

The measured signal strength can be used to evaluate the product \( \sigma \cdot B_{H \rightarrow WW^*} \) for Higgs boson production at a mass \( m_H = 125.36 \) GeV, as well as for the individual ggF and VBF production modes. The central value is simply the product of \( \mu \) and the predicted cross section used to define it. The uncertainties are similarly scaled, except for the theoretical uncertainties related to the total production yield, which do not apply to this measurement. These are the QCD scale and PDF uncertainties on the total cross sections, and the uncertainty on the branching fraction for \( H \rightarrow WW^* \), as described in Sec. V. In practice, the corresponding nuisance parameters are fixed to their nominal values in the fit, effectively removing these uncertainties from consideration. Inclusive cross-section measurements are performed for ggF and VBF production. The cross section is also measured for ggF production in...
defined fiducial volumes; this approach minimizes the impact of theoretical uncertainties.

1. Inclusive cross sections

Inclusive cross sections are evaluated at both 7 and 8 TeV for the ggF production process and at 7 TeV for the VBF production process. The 7 TeV VBF cross section is not measured because of the large statistical uncertainty. The signal strengths used for ggF and VBF are determined through a simultaneous fit to all categories as described in Sec. IX C. The small VH contribution, corresponding to a 9.9\% uncertainty, is neglected, and its expected fractional yield is added linearly to the total uncertainty. The 7 TeV signal strength $\mu_{\text{ggF}}^7$ and 8 TeV signal strengths $\mu_{\text{ggF}}^8$ and $\mu_{\text{VBF}}^8$ are

$$
\begin{align*}
\mu_{\text{ggF}}^7 & = 0.57_{-0.51}^{+0.52} & \mu_{\text{ggF}}^8 & = 1.09_{-0.17}^{+0.20} & \mu_{\text{VBF}}^8 & = 1.45_{-0.44}^{+0.48} \\
\mu_{\text{ggF}}^7 & = 0.36_{-0.34}^{+0.36} & \mu_{\text{ggF}}^8 & = 0.19_{-0.17}^{+0.20} & \mu_{\text{VBF}}^8 & = 0.38_{-0.24}^{+0.38} \\
\mu_{\text{ggF}}^7 & = 0.14_{-0.04}^{+0.14} & \mu_{\text{ggF}}^8 & = 0.14_{-0.09}^{+0.17} & \mu_{\text{VBF}}^8 & = 0.11_{-0.06}^{+0.24} \\
\end{align*}
$$

where (sig) indicates the systematic uncertainties on the total signal yield for the measured process, which do not affect the cross-section measurement. The effect of uncertainties on the signal yield for other production modes is included in the systematic uncertainties. In terms of the measured signal strength, the inclusive cross section is defined as

$$
\sigma \cdot B_{H \rightarrow WW} = \frac{(N_{\text{sig}})_{\text{obs}}}{A \cdot C \cdot B_{\text{WW}} \cdot \text{eff} \cdot \int L dt} = \hat{\mu} \cdot (\sigma \cdot B_{H \rightarrow WW})_{\text{exp}}.
$$

(20)

In this equation, $A$ is the kinematic and geometric acceptance, and $C$ is the ratio of the number of measured events to the number of events produced in the fiducial phase space of the detector. The product $A \cdot C$ is the total acceptance for reconstructed events. The cross sections are measured using the last line of the equation, and the results are

$$
\begin{align*}
\sigma_{\text{ggF}}^7 \cdot B_{H \rightarrow WW} & = 2.0 \pm 1.7_{-1.1}^{+1.2} \quad = 2.0 \pm 2.1 \text{ pb}
\sigma_{\text{ggF}}^8 \cdot B_{H \rightarrow WW} & = 4.6 \pm 0.9_{-0.7}^{+0.8} \quad = 4.6 \pm 1.4 \text{ pb}
\sigma_{\text{VBF}}^8 \cdot B_{H \rightarrow WW} & = 0.51 \pm 0.17_{-0.08}^{+0.13} \quad = 0.51 \pm 0.22 \text{ pb,}
\end{align*}
$$

(21)

The predicted cross-section values are 3.3 ± 0.4 pb, 4.2 ± 0.5 pb, and 0.35 ± 0.02 pb, respectively.

These are derived as described in Sec. V, and the acceptance is evaluated using the standard signal MC samples.

2. Fiducial cross sections

Fiducial cross-section measurements enable comparisons to theoretical predictions with minimal assumptions about the kinematics of the signal and possible associated jets in the event. The cross sections described here are for events produced within a fiducial volume closely corresponding to a ggF signal region. The fiducial volume is defined using generator-level kinematic information, as specified in Table XXVIII. In particular, the total $p_T$ of the

TABLE XXVIII. Fiducial volume definitions for fiducial cross sections. The selection is made using only $e\mu$ events. Events in which one or both $W$ bosons decay to $\tau\tau$ are excluded from the fiducial volume, but are present in the reconstructed volume. Energy-related quantities are in GeV.

<table>
<thead>
<tr>
<th>Type</th>
<th>$n_j = 0$</th>
<th>$n_j = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_T^\ell &gt; 22$</td>
<td>$p_T^\ell &gt; 10$</td>
</tr>
<tr>
<td>Preselection</td>
<td>$p_T^\ell &gt; 10$</td>
<td>opposite-charge $\ell$</td>
</tr>
<tr>
<td></td>
<td>$m_{\ell\ell} &gt; 10$</td>
<td>$m_{\ell\ell} &gt; 50$</td>
</tr>
<tr>
<td>$n_j$ dependent</td>
<td>$m_{\ell\ell} &gt; 66$</td>
<td>$m_{\ell\ell} &lt; 55$</td>
</tr>
<tr>
<td></td>
<td>$\Delta \phi_{\ell\ell} &lt; 1.8$</td>
<td>$\Delta \phi_{\ell\ell} &lt; 1.8$</td>
</tr>
</tbody>
</table>

012006-60
neutrino system ($p_{T}^{\nu}$) replaces the $p_{T}^{\text{miss}}$, and each lepton’s $p_{T}$ is replaced by the generated lepton $p_{T}$, where the lepton four-momentum is corrected by adding the four-momenta of all photons within a cone of size $\Delta R = 0.1$ to account for energy loss through QED final-state radiation. These quantities are used to compute $m_{T}$. Jets are defined at hadron level, i.e., after parton showering and hadronization but before detector simulation. To minimize dependence on the signal model, and therefore the theoretical uncertainties, only $e\mu$ events in the $n_j \leq 1$ categories are used. Also, only the 8 TeV data sample is used for these measurements.

The measured fiducial cross section is defined as

$$\sigma_{\text{fid}} = \frac{(N_{\text{sig}})_{\text{obs}}}{C} \cdot \frac{1}{\int L \, dt} = \hat{\mu} \cdot (\sigma \cdot B_{H\rightarrow WW^{*}\rightarrow e\mu})_{\text{exp}} \cdot A, \quad (22)$$

with the multiplicative factor $A$ being the sole difference with respect to the inclusive cross-section calculation. The measured fiducial cross section is not affected by the theoretical uncertainties on the total signal yield nor by the theoretical uncertainties on the signal acceptance. The total uncertainty is reduced compared to the value for the inclusive cross section because the measured signal yield is not extrapolated to the total phase space.

The correction factors for $n_j = 0$ and $n_j = 1$ events, $C_{0j}^{\text{ggF}}$ and $C_{1j}^{\text{ggF}}$, are evaluated using the standard signal MC sample. The reconstructed events include leptons from $\tau$ decays, but for simplicity, the fiducial volume is defined without these contributions. According to the simulation, the fraction of measured signal events within the fiducial volume is 85% for $n_j = 0$ and 63% for $n_j = 1$.

The values of the correction factors are

$$C_{0j}^{\text{ggF}} = 0.507 \pm 0.027, \quad C_{1j}^{\text{ggF}} = 0.506 \pm 0.022. \quad (23)$$

The experimental systematic uncertainty is approximately 5%. Remaining theoretical uncertainties on the $C_{0j}^{\text{ggF}}$ values were computed by comparing the ggF predictions of POWHEG + HERWIG, POWHEG + PYTHIA8, and POWHEG + PYTHIA6, and are found to be approximately 2% and are neglected. The acceptance of the fiducial volume is

$$A_{0j}^{\text{ggF}} = 0.206 \pm 0.030, \quad A_{1j}^{\text{ggF}} = 0.075 \pm 0.017. \quad (24)$$

The uncertainties on the acceptance are purely theoretical in origin and the largest contributions are from the effect of the QCD scale on the jet multiplicity requirements.

The cross-section values are computed by fitting the $\mu$ values in the $n_j = 0$ and $n_j = 1$ categories. The VBF contribution is subtracted assuming the expected yield from the SM instead of using the simultaneous fit to the VBF signal regions as is done for the inclusive cross sections. The non-negligible ggF yield in the VBF categories would require an assumption on the ggF acceptance for different jet multiplicities, whereas the fiducial cross-section measurement is intended to avoid this type of assumption. The effect of the theoretical uncertainties on the VBF signal yield is included in the systematic uncertainties on the cross sections. The obtained signal strengths are

$$\mu_{0j, e\mu}^{\text{ggF}} = 1.39 \pm 0.27 \pm 0.21 \pm 0.27 \pm 0.19 \pm 0.17 \pm 0.19 \pm 0.17, \quad \mu_{1j, e\mu}^{\text{ggF}} = 1.14 \pm 0.42 \pm 0.27 \pm 0.42 \pm 0.41 \pm 0.26 \pm 0.42 \pm 0.17 \pm 0.17. \quad (25)$$

where (sig) indicates the systematic uncertainties on the signal yield and acceptance, which do not apply to the fiducial cross-section measurements. The corresponding cross sections, evaluated at $m_{H} = 125.36$ GeV and using the 8 TeV data, are

$$\sigma_{\text{fid}, 0j}^{\text{ggF}} = 27.6 \pm 5.4 \pm 4.1 \pm 3.9 = 27.6 \pm 6.6 \text{ fb}, \quad \sigma_{\text{fid}, 1j}^{\text{ggF}} = 8.3 \pm 3.1 \pm 2.0 \pm 1.9 = 8.3 \pm 3.5 \text{ fb.} \quad (26)$$

The predicted values are 19.9 ± 3.3 fb and 7.3 ± 1.8 fb, respectively.

X. CONCLUSIONS

An observation of the decay $H \rightarrow WW^{*} \rightarrow \ell\ell\nu\nu$ with a significance of 6.1 standard deviations is achieved by an analysis of ATLAS data corresponding to 25 fb$^{-1}$ of integrated luminosity from $\sqrt{s} = 7$ and 8 TeV $pp$ collisions produced by the Large Hadron Collider at CERN. This observation confirms the predicted decay of the Higgs boson to $W$ bosons, at a rate consistent with that given by the Standard Model. The SM predictions are additionally supported by evidence for VBF production in this channel, with an observed significance of 3.2 standard deviations.

For a Higgs boson with a mass of 125.36 GeV, the ratios of the measured cross sections to those predicted by the Standard Model are consistent with unity for both gluon-fusion and vector-boson-fusion production:

$$\mu = 1.09 \pm 0.23 \pm 0.21, \quad \mu_{\text{ggF}} = 1.02 \pm 0.29 \pm 0.26, \quad \mu_{\text{VBF}} = 1.27 \pm 0.53 \pm 0.45. \quad (27)$$

The measurement uncertainties are reduced by 30% relative to the prior ATLAS $H \rightarrow WW^{*} \rightarrow \ell\ell\nu\nu$ measurements due to improved analysis techniques. The corresponding cross section times branching fraction values are
These total cross sections, as well as the fiducial cross sections measured in the exclusive \( n_j = 0 \) and \( n_j = 1 \) categories, allow future comparisons to the more precise cross section calculations currently under development.

The analysis strategies described in this paper set the stage for more precise measurements using future collisions at the LHC. The larger data sets will significantly reduce statistical uncertainties; further modeling and analysis improvements will be required to reduce the leading systematic uncertainties. Future precise measurements of the \( H \rightarrow WW^* \) decay will provide more stringent tests of the detailed SM predictions of the Higgs boson properties.

ACKNOWLEDGMENTS

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNSRC and Lundbeck Foundation, Denmark; EPLANET, ERC and NSRF, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, DFS, HGF, MPG and AvH Foundation, Germany; GSRT and NSRF, Greece; ISF, MINERVA, GIF, DIP and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; BRF and RCN, Norway; MNiSW, Poland; GRICES and FCT, Portugal; MERSYS (MECTS), Romania; MES of Russia and ROSATOM, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

APPENDIX A: STATISTICAL TREATMENT DETAILS

1. Binning of fit variables

The \( m_T \) distribution is used in the likelihood fit for the ggF-enriched \( n_j \) samples (see Sec. VII). Figure 43 shows an example of the binned \( m_T \) distribution in the most sensitive kinematic region of \( n_j = 0 \) and \( e\mu \) lepton-flavor category. The optimization procedure for the widths was discussed in Sec. VII A. Table XXIX gives the details of the binning for every kinematic region. The \( m_T \) range between the bin 1 (around 80 GeV) and the last bin (around 120 GeV) is binned in variable widths. For kinematic regions in the \( n_j = 0 \) category, the variable widths are approximately 6 GeV; for \( n_j = 1 \), the widths are approximately 10 GeV. For both samples, the rms of the bin widths from the mean bin width is approximately 1 GeV. Lastly, the ggF-enriched \( n_j \geq 2 \) and the cross-check VBF-enriched \( n_j \geq 2 \) categories use the same set of fixed \( m_T \) bin boundaries with bins of variable width.

2. Drell-Yan estimate in ee/\( \mu\mu \) for \( n_j \leq 1 \)

The details of the treatment for the Drell-Yan estimate for the ee/\( \mu\mu \) category in the \( n_j \leq 1 \) sample are described.

The method uses additional control regions to constrain the parameters corresponding to the selection efficiencies of contributing processes categorized into “DY” and “non-DY”; the latter includes the signal events. The variable \( f_{\text{recoil}} \) is used to separate the two categories, and to divide the sample into “pass” and “fail” subsamples. In the ee/\( \mu\mu \) categories, the pass samples are enriched in non-DY events and, conversely, the fail samples are enriched in DY events. The residual cross-contamination is estimated using additional control regions.

\[ \sigma_{\gamma \gamma F} \cdot B_{H\rightarrow WW^*} = 2.0 \pm^{+2.1}_{-2.0} \text{ pb}, \]
\[ \sigma_{W W} \cdot B_{H\rightarrow WW^*} = 4.6 \pm^{+1.2}_{-1.1} \text{ pb}, \]
\[ \sigma_{VBF} \cdot B_{H\rightarrow WW^*} = 0.51 \pm^{+0.22}_{-0.17} \text{ pb}. \]
The non-DY events with neutrino final states, or “real missing transverse momentum,” contaminate both the Z CR and the SR, and the two corresponding \( f_{\text{recoil}} \) selection efficiencies, \( \epsilon'_{\text{non-DY}} \) and \( \epsilon_{\text{non-DY}} \), are evaluated separately. The non-DY efficiency in the Z CR, \( \epsilon'_{\text{non-DY}} \), is evaluated using the Z CR selection except with the \( \mu \mu \) sample, which is pure in non-DY events. The SR efficiency \( \epsilon_{\text{non-DY}} \) is evaluated using the \( ee/\mu\mu \) SR selection (described in Sec. VI E 2) applied to an \( \mu \mu \) sample.

The fit CR part of the likelihood function [Eq. (11)] contains two Poisson functions that represent events—in the Z mass window in the \( ee/\mu\mu \) category—that pass or fail the \( f_{\text{recoil}} \) selection:

Of particular interest is the data-derived efficiency of the \( f_{\text{recoil}} \) selection for the DY and non-DY events. The efficiency of the applied \( f_{\text{recoil}} \) selection on DY events (on \( \epsilon_{\text{DY}} \)) is obtained from the \( ee/\mu\mu \) sample in the Z peak (in the Z CR), defined by the dilepton mass range \( |m_{ee} - m_Z| < 15 \text{ GeV} \). Events in the Z CR are relatively pure in DY. The \( \epsilon_{\text{DY}} \) estimates the efficiency of the selection due to neutrinoless events with missing transverse momentum due to misreconstruction, or “fake missing transverse momentum.” The same parameter appears in two terms, one for the Z CR and the other for the signal region, each composed of two Poisson functions.

### Table XXIX.

<table>
<thead>
<tr>
<th>Sample ( \ell_2 )</th>
<th>( m_{ee} )</th>
<th>( p_T^2 )</th>
<th>( n_{\text{bins}} )</th>
<th>Category</th>
<th>Bin left edge</th>
<th>Bin widths ( w_b ) for bin ( b )</th>
<th>Mean width, rms of deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_{j} = 0 )</td>
<td>( 10)</td>
<td>( 10)</td>
<td>( 15)</td>
<td>( 10)</td>
<td>74.5</td>
<td>118.2</td>
<td>( 5.9 )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 10)</td>
<td>( 10)</td>
<td>( 15)</td>
<td>( 10)</td>
<td>81.6</td>
<td>122.1</td>
<td>( 6.3 )</td>
</tr>
<tr>
<td>( 20)</td>
<td>( \infty)</td>
<td>( 10)</td>
<td>93.7</td>
<td>133.7</td>
<td>( 6.3 )</td>
<td>4.6</td>
<td>3.9</td>
</tr>
<tr>
<td>( 20)</td>
<td>( \infty)</td>
<td>( 10)</td>
<td>84.1</td>
<td>124.7</td>
<td>( 6.4 )</td>
<td>4.6</td>
<td>4.4</td>
</tr>
<tr>
<td>( 15)</td>
<td>( 20)</td>
<td>( 10)</td>
<td>86.3</td>
<td>125.8</td>
<td>( 6.0 )</td>
<td>4.7</td>
<td>4.4</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( \infty)</td>
<td>( 10)</td>
<td>93.2</td>
<td>135.4</td>
<td>( 7.0 )</td>
<td>4.8</td>
<td>4.2</td>
</tr>
<tr>
<td>( 10)</td>
<td>( 30)</td>
<td>( 10)</td>
<td>76.7</td>
<td>118.0</td>
<td>( 5.8 )</td>
<td>4.5</td>
<td>4.2</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 10)</td>
<td>( 5)</td>
<td>80.8</td>
<td>121.4</td>
<td>( 5.9 )</td>
<td>4.8</td>
<td>4.4</td>
</tr>
<tr>
<td>( 10)</td>
<td>( 20)</td>
<td>( 10)</td>
<td>93.1</td>
<td>133.6</td>
<td>( 6.7 )</td>
<td>4.9</td>
<td>4.0</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 30)</td>
<td>( 10)</td>
<td>84.9</td>
<td>125.7</td>
<td>( 6.0 )</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 15)</td>
<td>( 20)</td>
<td>85.0</td>
<td>125.2</td>
<td>( 6.6 )</td>
<td>4.8</td>
<td>4.1</td>
</tr>
<tr>
<td>( 10)</td>
<td>( 20)</td>
<td>( 10)</td>
<td>93.5</td>
<td>135.8</td>
<td>( 6.8 )</td>
<td>4.9</td>
<td>4.2</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 15)</td>
<td>( 30)</td>
<td>( 10)</td>
<td>10.1</td>
<td>122.8</td>
<td>( 4.9 )</td>
<td>4.0</td>
</tr>
</tbody>
</table>

### Table parameters:
- \( \mu \mu \) bin 2 left edge
- Bin widths \( w_b \) for bin \( b \)
- Mean width, rms of deviation
- Plot of \( \bar{\omega} \pm \text{rms} \)
\[ f(N_{\text{pass}}^{\text{ZCR}} | \beta_\text{DY}, \epsilon_\text{DY}, B_{\text{DY}}^{\text{ZCR}} + \epsilon_{\text{non-DY}} \cdot B_{\text{non-DY}}^{\text{ZCR}}), \]
\[ f(N_{\text{fail}}^{\text{ZCR}} | \beta_\text{DY}, (1 - \epsilon_\text{DY}) \cdot B_{\text{DY}}^{\text{ZCR}} + (1 - \epsilon_{\text{non-DY}}) \cdot B_{\text{non-DY}}^{\text{ZCR}}), \tag{A1} \]

where \( N \) is the observed number of events and \( B \) the background estimate without applying an \( f_{\text{recoil}} \) selection. The superscript denotes the Z CR mass window; the subscript pass (fail) denotes the sample of events that pass (fail) the \( f_{\text{recoil}} \) selection; and the subscripts DY (non-DY) denotes background estimates for the Drell-Yan (all except Drell-Yan) processes. The non-DY estimate, \( B_{\text{non-DY}}^{\text{ZCR}} \), is a sum of all contributing processes listed in Table I; normalization factors, such as \( \beta_{\text{WW}} \), that are described in Sec. VI are implicitly applied to the corresponding contributions. The Drell-Yan estimate is normalized explicitly by a common normalization factor \( \beta_{\text{DY}}^{\text{eff}} \) applied to both the passing and failing subsamples of the \( Z \) peak.

The \( \epsilon_{\text{non-DY}} \) parameter above is determined using events in the \( \mu \mu \) category. The corresponding Poisson functions are included in the likelihood:

\[ f(N_{\text{pass}}^{\text{ZCR},\epsilon\mu} | \epsilon_{\text{non-DY}} \cdot B_{\text{non-DY}}^{\text{ZCR},\epsilon\mu}), \]
\[ f(N_{\text{fail}}^{\text{ZCR},\epsilon\mu} | (1 - \epsilon_{\text{non-DY}}) \cdot B_{\text{non-DY}}^{\text{ZCR},\epsilon\mu}), \tag{A2} \]

where the \( \epsilon\mu \) in the superscript denotes the Z CR mass window for events in the \( \mu \mu \) category; all other notation follows the convention for Eq. (A1). The DY contamination in this region is implicitly subtracted.

The SR part of the likelihood also contains two Poisson functions—using the same \( \epsilon_\mu \) above, but a different \( \beta_\text{DY} \) and \( \epsilon_{\text{non-DY}} \) corresponding to the SR—is

\[ f(N_{\text{pass}}^{\text{SR}} | \beta_\text{DY} \cdot \epsilon_\text{DY} \cdot B_{\text{DY}}^{\text{SR}} + \epsilon_{\text{non-DY}} \cdot B_{\text{non-DY}}^{\text{SR}}), \]
\[ f(N_{\text{fail}}^{\text{SR}} | \beta_\text{DY} \cdot (1 - \epsilon_\text{DY}) \cdot B_{\text{DY}}^{\text{SR}} + (1 - \epsilon_{\text{non-DY}}) \cdot B_{\text{non-DY}}^{\text{SR}}), \tag{A3} \]

where SR denotes the signal region selection and \( \beta_\text{DY} \) is the common normalization factor for the Drell-Yan estimate for the pass and fail subsamples.

The parameter \( \epsilon_{\text{non-DY}} \) is constrained following the same strategy as Eq. (A2) with

\[ f(N_{\text{pass}}^{\text{SR},\epsilon\mu} | \epsilon_{\text{non-DY}} \cdot B_{\text{non-DY}}^{\text{SR},\epsilon\mu}), \]
\[ f(N_{\text{fail}}^{\text{SR},\epsilon\mu} | (1 - \epsilon_{\text{non-DY}}) \cdot B_{\text{non-DY}}^{\text{SR},\epsilon\mu}), \tag{A4} \]

where the \( \epsilon\mu \) in the superscript denotes the ee/\( \mu \mu \) SR selection (including the one on \( f_{\text{recoil}} \)) applied to events in the \( \mu \mu \) category. As noted before, the DY contamination in this region is implicitly subtracted.

3. Top-quark estimate for \( n_j = 1 \)

The details of the in situ treatment for the \( b \)-tagging efficiency for the top-quark estimate for \( n_j = 1 \) category are described.

The method uses two control regions within the \( n_j = 2 \) sample: those with one and two \( b \)-jets. These CRs constrain the normalization parameter for the \( b \)-tagging efficiency of top-quark events (\( \beta_{b-tag} \)) and for the top-quark cross section in these regions (\( \beta_{\text{top}} \)).

The Poisson terms for the control regions are

\[ f(N_{\text{2j}}^{\text{DY}} | \beta_{\text{top}} \cdot \beta_{b-tag} \cdot B_{\text{top}}^{\text{2j}} + B_{\text{other}}), \]
\[ f(N_{\text{1j}}^{\text{DY}} | \beta_{\text{top}} \cdot B_{\text{top}}^{\text{1j}} + \beta_{b-tag} \cdot (1 - \beta_{b-tag}) \cdot B_{\text{top}} + B_{\text{other}}), \tag{A5} \]

where \( N_{\text{1j}}^{\text{DY}} (N_{\text{2j}}^{\text{DY}}) \) corresponds to the number of observed events with one (two) \( b \)-jets; \( B_{\text{top}}^{\text{1j}} (B_{\text{top}}^{\text{2j}}) \) is the corresponding top-quark estimates from MC samples; and \( B_{\text{other}} \) are the rest of the processes contributing to the sample.

The parameter \( \beta_{\text{top}} \) enters only in the above terms, while \( \beta_{b-tag} \) is applied to other regions. In the top-quark CR, one factor of \( \beta_{\text{top}} \) is applied to the expected top-quark yield. In the SR and the WW CR, the treatment is of the same form as the second line of Eq. (A5) applied to the \( n_j = 1 \) sample, i.e., the estimated top-quark background is \( B_{\text{top}}^{\text{1j}} + (1 - \beta_{b-tag}) \cdot B_{\text{top}}^{\text{1j}} \).

In summary, the difference between the observed and the expected \( b \)-tagging efficiency corrects the number of estimated untagged events in the SR.

APPENDIX B: BDT PERFORMANCE

Section IV C motivated the choice of variables used in the \( n_j \geq 2 \) VBF-enriched category based on their effectiveness in the cross-check analysis. Many of the variables exploit the VBF topology with two forward jets and no activity in the central region. The main analysis in this category is based on the multivariate technique that uses those variables as inputs to the training of the BDT. The training is optimized on the simulated VBF signal production and it treats simulated ggF production as a background. Figures 44 and 45 show the distributions of the input variables in the \( \mu \mu \) and \( ee/\mu \mu \) samples, respectively. The comparison is based only on MC simulation and it shows the separation between the VBF signal and the background processes, motivating the use of the chosen variables.

The \( O_{\text{BDT}} \) distributions are shown in Fig. 46. The lowest \( O_{\text{BDT}} \) score is assigned to the events that are classified as background, and the highest score selects the VBF signal events. This separation can be seen in these distributions. The final binning configuration is four bins with boundaries
FIG. 44 (color online). Distributions of the variables used as inputs to the training of the BDT in the $e\mu$ sample in the 8 TeV data analysis. The variables are shown after the common preselection and the additional selection requirements in the $n_{j} \geq 2$ VBF-enriched category, and they include: $m_{ll}$, $\Delta \phi_{ll}$, $m_{T}$, and $\Delta y_{jj}$ (top two rows); $m_{jj}$, $p_{T}^{\text{sum}}$, $\Sigma C_{l}$, and $\Sigma m_{ej}$ (bottom two rows). The distributions show the separation between the VBF signal and background processes (ggF signal production is treated as such). The VBF signal is scaled by 50 to enhance the differences in the shapes of the input variable distributions. The SM Higgs boson is shown at $m_{H} = 125$ GeV. The uncertainties on the background prediction are only due to MC sample size.
FIG. 45 (color online). Distributions of the variables used as inputs to the training of the BDT in the ee/μμ sample in the 8 TeV data analysis. The variables are shown after the common preselection and the additional selection requirements in the $n_\ell \geq 2$ VBF-enriched category, and they include $m_{ll}$, $\Delta \phi_{ll}$, $m_T$, and $\Delta y_{jj}$ (top two rows); $m_{jj}$, $p_T^{\text{sum}}$ , $\Sigma C_l$, and $\Sigma m_{lj}$ (bottom two rows). The distributions show the separation between the VBF signal and background processes (ggF signal production is treated as such). The VBF signal is scaled by 50 to enhance the differences in the shapes of the input variable distributions. The SM Higgs boson is shown at $m_H = 125$ GeV. The uncertainties on the background prediction are only due to MC sample size.
at $[-1, -0.48, 0.3, 0.78, 1]$, and with bin numbering from 0 to 3. The background estimation and the signal extraction is then performed in bins of $O_{\text{BDT}}$. Figure 47 shows the data-to-MC comparison of the input variables in the three highest $O_{\text{BDT}}$ bins. Good agreement is observed in all the distributions. The event properties of the observed events in the highest BDT bin in the $n_j \geq 2$ VBF-enriched category are shown in Table XXX.
FIG. 47 (color online). Comparisons of the observed and expected distributions of the variables used as inputs to the training of the BDT in the $\mu\mu$ samples in the 8 TeV data analysis. The variables are shown after the common preselection and the additional selection requirements in the $n_{\ell}\geq2$ VBF-enriched category, and after the BDT classification selecting the final signal region ($O_{\text{BDT}}>-0.48$). The variables shown are $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$, $m_T$, and $p_T^{\text{sum}}$ (top two rows); $\Delta y_{jj}$, $m_{jj}$, $\Sigma C_1$, and $\Sigma m_j$ (bottom two rows). The SM Higgs boson is shown at $m_H=125$ GeV. Both the statistical and systematic uncertainties are included.
TABLE XXX. Event properties of the Higgs candidates in BDT bin 3 (\(O_{BDT} > 0.78\)) of the \(n_{j} \geq 2\) VBF-enriched category in the 8 TeV data analysis. The energy-related quantities are in GeV.

<table>
<thead>
<tr>
<th>Sample</th>
<th>(O_{BDT})</th>
<th>(m_T)</th>
<th>(m_{\ell\ell})</th>
<th>(\Delta \phi_{\ell\ell})</th>
<th>(\Sigma C_{\ell})</th>
<th>(m_{jj})</th>
<th>(\Delta y_{jj})</th>
<th>(\Sigma m_{\ell\ell})</th>
<th>(p_{T}^{\text{miss}})</th>
<th>(p_{T}^{\ell})</th>
<th>(\eta_{\ell})</th>
<th>(\eta_{\ell\ell})</th>
<th>(p_{T})</th>
<th>(\eta_{T})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e\mu) sample</td>
<td>0.92</td>
<td>113</td>
<td>19.8</td>
<td>0.26</td>
<td>0.25</td>
<td>1290</td>
<td>5.30</td>
<td>810</td>
<td>3.2</td>
<td>50</td>
<td>15</td>
<td>0.6</td>
<td>−0.1</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>0.88</td>
<td>102</td>
<td>27.2</td>
<td>0.62</td>
<td>0.83</td>
<td>530</td>
<td>3.75</td>
<td>540</td>
<td>2.5</td>
<td>73</td>
<td>20</td>
<td>1.6</td>
<td>−0.5</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>87</td>
<td>19.9</td>
<td>0.02</td>
<td>1.01</td>
<td>1770</td>
<td>7.55</td>
<td>1510</td>
<td>4.8</td>
<td>84</td>
<td>21</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>105</td>
<td>26.5</td>
<td>0.64</td>
<td>0.86</td>
<td>1500</td>
<td>5.58</td>
<td>1250</td>
<td>7.4</td>
<td>48</td>
<td>20</td>
<td>0.5</td>
<td>0.4</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>43</td>
<td>31.7</td>
<td>1.08</td>
<td>0.05</td>
<td>850</td>
<td>5.44</td>
<td>710</td>
<td>2.2</td>
<td>29</td>
<td>18</td>
<td>−1.5</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>84</td>
<td>37.2</td>
<td>0.18</td>
<td>0.76</td>
<td>810</td>
<td>5.26</td>
<td>660</td>
<td>4.8</td>
<td>34</td>
<td>23</td>
<td>−0.5</td>
<td>−0.2</td>
<td>101</td>
</tr>
<tr>
<td>(e/\mu) sample</td>
<td>0.92</td>
<td>66</td>
<td>38.9</td>
<td>0.18</td>
<td>0.33</td>
<td>1370</td>
<td>5.15</td>
<td>1060</td>
<td>2.2</td>
<td>147</td>
<td>44</td>
<td>1.4</td>
<td>1.1</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>102</td>
<td>27.4</td>
<td>0.65</td>
<td>0.24</td>
<td>1260</td>
<td>5.97</td>
<td>780</td>
<td>2.5</td>
<td>32</td>
<td>29</td>
<td>1.3</td>
<td>0.8</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>0.89</td>
<td>81</td>
<td>26.3</td>
<td>0.08</td>
<td>0.57</td>
<td>670</td>
<td>3.90</td>
<td>610</td>
<td>6.6</td>
<td>54</td>
<td>25</td>
<td>−1.7</td>
<td>−1.3</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>0.88</td>
<td>118</td>
<td>57.6</td>
<td>2.31</td>
<td>0.28</td>
<td>700</td>
<td>3.90</td>
<td>650</td>
<td>2.9</td>
<td>46</td>
<td>15</td>
<td>−0.4</td>
<td>−2.0</td>
<td>123</td>
</tr>
</tbody>
</table>


OBSERVATION AND MEASUREMENT OF HIGGS BOSON ...


OBSERVATION AND MEASUREMENT OF HIGGS BOSON …


(ATLAS Collaboration)

1Department of Physics, University of Adelaide, Adelaide, Australia
2Physics Department, SUNY Albany, Albany NY, United States of America
3Department of Physics, University of Alberta, Edmonton AB, Canada
4aDepartment of Physics, Ankara University, Ankara, Turkey
4bDepartment of Physics, Gazi University, Ankara, Turkey
4cIstanbul Aydin University, Istanbul, Turkey
4dDivision of Physics, TOBB University of Economics and Technology, Ankara, Turkey
5LAPP, CNRS/IN2P3 and Université de Savoie, Annecy-le-Vieux, France
6High Energy Physics Division, Argonne National Laboratory, Argonne IL, United States of America
7Department of Physics, University of Arizona, Tucson AZ, United States of America
8Department of Physics, The University of Texas at Arlington, Arlington TX, United States of America
9Physics Department, University of Athens, Athens, Greece
10Physics Department, National Technical University of Athens, Zografou, Greece
11Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
12Institut de Física d’Altes Energies and Departament de Física de la Universitat Autònoma de Barcelona, Barcelona, Spain
13Institute of Physics, University of Belgrade, Belgrade, Serbia
13bVinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
14Department for Physics and Technology, University of Bergen, Bergen, Norway
15Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, United States of America
16Department of Physics, Humboldt University, Berlin, Germany
17Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
18School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
18aDepartment of Physics, Bogazici University, Istanbul, Turkey
19Department of Physics, Bogazici University, Istanbul, Turkey
19cDepartment of Physics Engineering, Gaziantep University, Gaziantep, Turkey
20aINFN Sezione di Bologna, Italy
20bDipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy
21Physikalisches Institut, Universität zu Köln, Cologne, Germany
22Department of Physics, Boston University, Boston MA, United States of America
23Department of Physics, Brandeis University, Waltham MA, United States of America
24Universidade Federal do Rio de Janeiro COPPE/EE/IF, Rio de Janeiro, Brazil
24aElectrical Circuits Department, Federal University of Juiz de Fora (UFJF), Juiz de Fora, Brazil
24bFederal University of Sao Joao del Rei (UFSJ), Sao Joao del Rei, Brazil
24cInstituto de Fisica, Universidade de Sao Paulo, Sao Paulo, Brazil
25Physics Department, Brookhaven National Laboratory, Upton NY, United States of America
26aNational Institute of Physics and Nuclear Engineering, Bucharest, Romania
26bNational Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj Napoca, Romania
26cUniversity Politehnica Bucharest, Bucharest, Romania
26dWest University in Timisoara, Timisoara, Romania
27Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
28 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
29 Department of Physics, Carleton University, Ottawa ON, Canada
30 CERN, Geneva, Switzerland
31 Enrico Fermi Institute, University of Chicago, Chicago IL, United States of America
32 Departamento de Física, Pontificia Universidad Católica de Chile, Santiago, Chile
33 Departamento de Física, Universidad Técnica Federico Santa María, Valparaiso, Chile
34 Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China
35 Department of Modern Physics, University of Science and Technology of China, Anhui, China
36 Department of Physics, Nanjing University, Jiangsu, China
37 School of Physics, Shandong University, Shandong, China
38 Physics Department, Tsinghua University, Beijing 100084, China
39 Laboratoire de Physique Corpusculaire, Clermont Université and Université Blaise Pascal and CNRS/IN2P3, Clermont-Ferrand, France
40 Nevis Laboratory, Columbia University, Irvington NY, United States of America
41 Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
42 INFN Gruppo Collegato di Cosenza, Laboratori Nazionali di Frascati, Italy
43 Physics Department, Shanghai Jiao Tong University, Shanghai, China
44 Dipartimento di Fisica, Università della Calabria, Rende, Italy
45 AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow, Poland
46 SUPA - School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
47 INFN Laboratori Nazionali di Frascati, Frascati, Italy
48 Fakultät für Mathematik und Physik, Albert-Ludwigs-Universität, Freiburg, Germany
49 Section de Physique, Université de Genève, Geneva, Switzerland
50 INFN Sezione di Genova, Italy
51 Marian Smoluchowski Institute of Physics, Jagiellonian University, Krakow, Poland
52 The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland
53 The High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
54 Il Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany
55 Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany
56 Department of Physics, Duke University, Durham NC, United States of America
57 SUPA - School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
58 Dipartimento di Fisica, Università di Genova, Genova, Italy
59 E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi, Georgia
60 High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
61 II Physikalisches Institut, Justus-Liebig-Universität Göttingen, Göttingen, Germany
62 Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany
63 Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble, France
64 Department of Physics, Hampton University, Hampton VA, United States of America
65 Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge MA, United States of America
66 Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
67 Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
68 ZITI Institut für technische Informatik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
69 Faculty of Applied Information Science, Hiroshima Institute of Technology, Hiroshima, Japan
70 Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
71 Department of Physics, The University of Hong Kong, Hong Kong, China
72 Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
73 Department of Physics, Indiana University, Bloomington IN, United States of America
74 Institut für Astro- und Teilchenphysik, Leopold-Franzens-Universität, Innsbruck, Austria
75 University of Iowa, Iowa City IA, United States of America
76 Department of Physics and Astronomy, Iowa State University, Ames IA, United States of America
77 KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
78 JINR Dubna, Dubna, Russia
79 Graduate School of Science, Kobe University, Kobe, Japan
80 Faculty of Science, Kyoto University, Kyoto, Japan
81 Kyoto University of Education, Kyoto, Japan
82 Joint Institute for Nuclear Research, JINR Dubna, Dubna, Russia
83 KEK, High Energy Accelerator Research Organization, Tsukuba, Japan