Study of High-Transverse-Momentum Higgs Boson Production in Association with a Vector Boson in the $qqbb$ Final State with the ATLAS Detector

Aad, G.; Abbott, B.; Abeling, K.; Abicht, N.J.; Abidi, S.H.; Aboulhorma, A; Abramowicz, H.; Abreu, H.; Abulaiti, Y.; Acharya, B.S.; Dam, Mogens; Camplani, Alessandra; Hansen, Peter Henrik; Hansen, Jørn Dines; Hansen, Jørgen Beck; Ignazzi, Rosanna; Petersen, Troels Christian; Wiglesworth, Graig; Xella, Stefania; ATLAS Collaboration

Published in:
Physical Review Letters

DOI:
10.1103/PhysRevLett.132.131802

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):

Download date: 01. maj. 2024
Study of High-Transverse-Momentum Higgs Boson Production in Association with a Vector Boson in the $qqbb$ Final State with the ATLAS Detector

G. Aad et al.*

(ATLAS Collaboration)

(Received 14 December 2023; accepted 26 February 2024; published 28 March 2024)

This Letter presents the first study of Higgs boson production in association with a vector boson ($V = W$ or $Z$) in the fully hadronic $qqbb$ final state using data recorded by the ATLAS detector at the LHC in proton-proton collisions at $\sqrt{s} = 13$ TeV and corresponding to an integrated luminosity of 137 fb$^{-1}$. The vector bosons and Higgs bosons are each reconstructed as large-radius jets and tagged using jet substructure techniques. Dedicated tagging algorithms exploiting $b$-tagging properties are used to identify jets consistent with Higgs bosons decaying into $bb$. Dominant backgrounds from multijet production are determined directly from the data, and a likelihood fit to the jet mass distribution of Higgs boson candidates is used to extract the number of signal events. The $VH$ production cross section is measured inclusively and differentially in several ranges of Higgs boson transverse momentum: 250–450, 450–650, and greater than 650 GeV. The inclusive signal yield relative to the standard model expectation is observed to be $\mu = 1.4^{+1.0}_{-0.9}$ and the corresponding cross section is $3.1 \pm 1.3(\text{stat})^{+1.8}_{-1.4}(\text{syst})$ pb.

DOI: 10.1103/PhysRevLett.132.131802

In the standard model (SM) of particle physics, the Brout-Englert-Higgs mechanism [1–4] spontaneously breaks electroweak symmetry. As a result, a physical Higgs boson emerges and both the $W$ and $Z$ gauge bosons and the fermions acquire mass. Since the discovery of the Higgs boson by the ATLAS and CMS experiments at the LHC [5,6], precision measurements of its properties have been a priority for both experiments. The Higgs boson decay into bottom-quark pairs ($H \rightarrow b\bar{b}$) has the largest branching fraction ($\sim 58\%$) [7] and has only recently been observed by exploiting associated production with a vector boson ($VH$, $V = W$, $Z$) [8,9]. In these analyses, the associated $W$ and $Z$ bosons were required to decay leptonically to provide both a means to trigger these events and an effective suppression of the large background from QCD production of $b$ quarks. Theoretical work has raised interest in testing for new physics contributions to associated $VH$ production at high momentum by exploiting $H \rightarrow b\bar{b}$ decays in a highly boosted event topology [10–12] because of its sensitivity to higher-order effective operators. On the experimental side, ATLAS and CMS have successfully developed and calibrated novel jet-substructure techniques to identify high-momentum $V$ or $H$ bosons that decay hadronically and are reconstructed as a single large-radius ("large-$R$") jet. Dedicated tagging algorithms exploiting $b$-tagging properties are used to identify jets consistent with Higgs bosons decaying into $bb$.

In this Letter, these experimental techniques are applied in a first study of associated $VH$ production at high Higgs boson transverse momentum ($p_T^H$) in the two large-$R$ jets topology. This approach has the potential to probe higher $p_T^H$ values than in the best current measurement in the semileptonic final state [13] because the $W/Z$ hadronic branching fractions are larger. The $VH$ signal is extracted using a likelihood fit to the invariant mass distribution of Higgs candidate large-$R$ jets in events also containing a $W$- or $Z$-candidate jet. The dominant multijet background is estimated directly from the data, while smaller backgrounds ($t\bar{t}$, $V + $ jets, and $VV$ production) are modeled using Monte Carlo (MC) simulation. The normalization of the peaking $Z + $ jets background (mostly from $Z \rightarrow b\bar{b}$ decays passing as Higgs candidates) is a free parameter in the final fit. The selected events are separated into several $p_T^H$ ranges and the $VH$ production cross section is measured both inclusively and as a function of $p_T^H$.

The analysis presented here uses the proton-proton collision data collected by the ATLAS detector [14–16] from 2015 to 2018, corresponding to an integrated luminosity of 137 fb$^{-1}$ at $\sqrt{s} = 13$ TeV. The ATLAS detector is a multipurpose particle detector with cylindrical geometry. It consists of an inner tracking detector surrounded by a superconducting solenoid, sampling electromagnetic and hadronic calorimeters, and a muon spectrometer with three toroidal superconducting magnets, providing a near $4\pi$ coverage in solid angle [17]. A two-level trigger

*Full author list given at the end of the Letter.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.
system [18] selects events for storage. An extensive software suite [19] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment. Collision events satisfy a number of requirements ensuring that the ATLAS detector was operating well while the data were recorded.

Simulated events were produced for a variety of processes. All generated events were passed through a full simulation of the ATLAS detector response [20] using GEANT4 [21]. The effects of multiple $p\bar{p}$ interactions in the same or neighboring bunch crossings (pileup) were included by overlaying events generated with PYTHIA8 [22]. Events were weighted such that the distribution of the average number of interactions per bunch crossing matches that observed in data.

Higgs boson events were generated at next-to-leading-order (NLO) accuracy via gluon-gluon fusion, vector-boson fusion, and in association with a vector boson ($VH$) using POWHEG BOX v2 [23–26] and the PDF4LHC15NLO parton distribution function (PDF) set [27]. The loop-induced $gg \rightarrow ZH$ process was generated separately at leading order (LO) with POWHEG BOX. In these cases, the events were interfaced with PYTHIA8 [22] utilizing the AZNLO tune [28] to incorporate parton shower and nonperturbative effects. The production of a Higgs boson in association with two top quarks ($t\bar{t}H$) was modeled similarly, using POWHEG BOX v2 with the NNPDF3.1NLO PDF set [29] and the A14 tune [30].

Other Higgs boson production mechanisms such as $tH$ and QCD-induced $bbH$ are not considered since their contribution is negligible. The simulated processes are normalized using the best available cross-section calculations [7]. In particular, the $pp \rightarrow VH$ cross section is calculated to next-to-next-to-leading order in QCD with NLO electroweak (EW) corrections, and the $gg \rightarrow ZH$ cross section is calculated at NLO and next-to-leading-logarithm accuracy [31–37].

The $q\bar{q}/qg \rightarrow ZH$ simulated events are normalized using the $pp \rightarrow ZH$ cross section (which combines the $q\bar{q}$, $qg$, and $gg$ contributions) after subtracting the $gg \rightarrow ZH$ cross section. Differential NLO EW corrections computed with HAWK [35] are applied to the $q\bar{q}/qg \rightarrow VH$ process as a function of vector-boson transverse momentum, $p_T^V$.

Simulated multijet events are only used to optimize the event selection since the multijet background is estimated directly from the data, as described later. They were generated using PYTHIA8 with LO matrix elements for dijet production. The NNPDF2.3LO PDF set [38] and A14 tune were used in all steps of the event generation. Samples of $V + \text{jets}$ events were generated with SHERPA2.2.8 [39] and the NNPDF3.1NLO PDF set at NLO accuracy for one additional parton and LO accuracy for up to four additional partons. All generated $V + \text{jets}$ samples have NLO QCD and LO EW accuracy. Additional NLO EW corrections are applied to the simulated events as weights. Diboson events ($VV = WW, WZ, ZZ$) were generated with SHERPA2.2.8 and the NNPDF3.1NLO PDF set at NLO accuracy for up to one additional parton and LO accuracy for up to three additional partons. The simulated $t\bar{t}$ events were generated at NLO using POWHEG BOX v2 with $h_{\text{damp}} = 1.5 m_{\text{top}}$ [40].

Weights are also applied to match the $t\bar{t}$ events measured in data in the same kinematic phase space [41]. Single-top production in the $Wt$ channel was modeled at NLO using POWHEG BOX v2 in the five-flavor scheme [42] and with the diagram removal scheme [40] to treat interference with $t\bar{t}$ production. The NNPDF3.1NLO PDF set was used for all top-quark samples. Except for Sherpa samples, the decay of $b$ and $c$ hadrons was performed using EvtGen1.6.0 [43].

Events are required to pass single large-$R$ jet triggers [18], with mass and transverse momentum exceeding specific thresholds. The event selection described below is designed to ensure full trigger efficiency. Events are also required to have a primary vertex with at least two associated tracks. The primary vertex is selected as the vertex with the largest $\Sigma p_T^2$, where the sum is over all of its associated tracks with $p_T > 0.5$ GeV [44]. In addition, noncollision backgrounds originating from calorimeter noise, beam halo interactions or cosmic rays are suppressed by rejecting events that contain any anti-$k_t$ calorimeter jet with $R = 0.4$ failing to satisfy a set of quality criteria [45].

Large-$R$ jets are used to identify both the $V$ boson and the Higgs boson. They are reconstructed with the anti-$k_t$ algorithm [46,47] with a radius parameter of $R = 1.0$, utilizing locally weighted topological cell clusters [48] for their constituents. At this stage, these jets are referred to as ungroomed jets. Jets are then trimmed [49] to remove energy deposits from pileup and the underlying event as well as soft and wide-angle radiation. The jet mass is reconstructed by combining the calorimeter and tracking measurements [50], where tracks are required to be associated with the primary vertex.

To aid in $H \rightarrow b\bar{b}$ tagging of large-$R$ jets, a separate collection of jets is built from tracks with the anti-$k_t$ algorithm using a $p_T$-dependent radius [51]. Variable-radius track jets are then matched to large-$R$ calorimeter jets via ghost association [52] and used as proxies for $b$ quarks associated with the Higgs candidate. All large-$R$ jets must satisfy $p_T > 200$ GeV and $|\eta| < 2.0$ to ensure they are contained within the tracking detector. Events with one or more isolated charged leptons (electrons or muons) are rejected. Electrons are identified by matching tracks to energy clusters in the electromagnetic calorimeter. They must have $p_T > 7$ GeV and $|\eta| < 2.5$, and satisfy the “loose” identification criterion defined in Ref. [53]. Muon identification relies on matching tracks in the inner detector to muon spectrometer tracks or track segments. Muons must have $p_T > 7$ GeV and $|\eta| < 2.5$, and satisfy the loose selection criterion [54].
The identification of $W$ and $Z$ bosons relies on several properties of the large-$R$ jets. The $D_2^{R=1}$ variable exploits the two-prong structure of the $W/Z \rightarrow q\bar{q}$ decays, absent in typical QCD jets [55,56]. The number of tracks ($N_{trk}$) linked to the ungroomed large-$R$ jets by ghost association is significantly higher for gluon-induced jets in background events than for quark-induced jets in signal events, due to the distinct energy scales involved and the different color factors for gluons and quarks. Requirements on the jet mass, $D_2^{R=1}$, and $N_{trk}$ are optimized as a function of jet $p_T$ such that a signal efficiency of 50% [57] is achieved in each jet-$p_T$ category, along with a multijet background rejection factor of 70–200. The $W$ boson tagging is calibrated with semileptonic $t\bar{t}$ data events and then extrapolated to the identification of $Z$ bosons [57]. For the high-$p_T$ category, an additional uncertainty is applied to the $Z$-boson tagging, based on differences between $W$-boson to $Z$-boson tagging efficiency extrapolations obtained with SHERPA and HERWIG.

Higgs candidates are identified by an $H \rightarrow b\bar{b}$ tagger [58] based on a neural-network algorithm that uses track and vertex information from the variable-radius track jets to discriminate between large-$R$ jets from Higgs boson decays into $b\bar{b}$ and jets from gluons, light quarks, or top quarks. A fixed 60% signal efficiency working point (WP) is used, where 60% refers to the average efficiency for selecting simulated Higgs bosons with $p_T > 250$ GeV [58]. The jet mass resolution for Higgs candidates is improved by applying a “muon-in-jet” correction to account for the energy carried by muons from semileptonic $b$-hadron decays [59]. The corrected jet mass of the Higgs boson candidate ($m^H$) is used to extract the signal.

Events are required to contain at least two selected large-$R$ jets. The leading (highest $p_T$) large-$R$ jet must have $p_T > 450$ GeV and mass above 50 GeV, to ensure full trigger efficiency. The second leading large-$R$ jet must have a mass above 40 GeV. At least one of those two jets must pass the $H \rightarrow b\bar{b}$ tagger requirements. If both jets satisfy those requirements, the one with a larger mass is selected as the Higgs candidate. The other jet must then satisfy the $W/Z$ tagging requirements. Furthermore, the Higgs candidate’s transverse momentum ($p_{T,1}^{H}$) must exceed 250 GeV, thereby defining the signal region (SR), and three exclusive ranges are chosen: $p_{T,1}^{H} \in [250, 450)$ GeV, $p_{T,1}^{H} \in [450, 650)$ GeV, and $p_{T,1}^{H} \geq 650$ GeV. In the SR, the $VH$ process dominates (85%), although other Higgs boson production processes contribute: $t\bar{t}H$ (8%), gluon-gluon fusion (6%), and vector-boson fusion (1.4%). The dominant background contribution comes from multijet production (90%), followed by $t\bar{t}$ (5%), $V +$ jets (3.6%), and diboson (0.7%) production. A data-driven method is used to estimate the background contribution from multijet production as well as that from $V +$ jets production in which the $W/Z$ boson is correctly tagged and the additional jet is a QCD jet misidentified as the Higgs candidate (both processes are referred to as “multijet”). Other background processes are modeled by simulation, including $V +$ jets production in which the $W/Z$ boson is selected as the Higgs candidate and a QCD jet passes the $V$ tagging requirements.

The multijet background estimate is extracted from a control region (CR) where events pass all SR requirements except the Higgs tagging requirement, since the $H \rightarrow b\bar{b}$ tagger does not rely on the jet mass as a discriminating variable. To account for differences in distribution shape and normalization between the SR and CR, a transfer factor (TF), which depends on both jet $p_T$ and jet mass ($m$), is used. The number of multijet events in the SR is thus derived from the number of multijet events in the CR ($N_{CR}^{multijet}$) as

$$N_{SR}^{multijet}(p_T, m) = TF(p_T, \rho) \times N_{CR}^{multijet}(p_T, m),$$

where $TF(p_T, \rho) = \sum_{k,l} \alpha_{kl} \rho^k \alpha_{l} \rho^l$, $\rho = \log(m^2/p_T^2)$, and $\alpha_{kl}$ are the polynomial coefficients for the $k$th order in $\rho$ and $l$th order in $p_T$. The $\alpha_{kl}$ coefficients are determined from a simultaneous fit to the data in the SR and CR across the whole jet mass range. To determine the order of the polynomial needed to fit the data, a Fisher $F$ test is performed. Based on its results, a first-order polynomial in both $\rho$ and $p_T$ is found to be sufficient to parametrize the transfer factor in the SR.

Three validation regions were defined to verify that the multijet background estimated using the TF method describes the observed background well. Events in these regions contain a $V$ jet passing the looser $W/Z$ tagging 80% efficiency WP [57] but failing either the $N_{trk}$, $D_2^{R=1}$, or mass requirement corresponding to the nominal 50% efficiency WP. The composition of events from background processes in these regions, as tested with simulated samples, is similar to that in the SR. The background modeling is found to be in good agreement with data within the statistical uncertainty in each of the validation regions.

A second multijet background estimation method is utilized to test the nominal TF method. A boosted decision tree (BDT), trained on an alternative set of data events that fail both the $V$ and Higgs boson tagging, is used to perform a kinematic reweighting to match the kinematic distributions in this alternative region to those in the SR. A description of this reweighting procedure is available in Ref. [60]. The two multijet background estimates agree within statistical uncertainties.

Systematic uncertainties arise from several different sources: the data-driven background estimate, the experimental reconstruction, and the theoretical prediction for the signal. Their impact is summarized in the Appendix. The uncertainty affecting the shape of the estimated multijet background is evaluated as the difference between the distributions from the BDT and TF methods. It is found to
be up to 10% for $p_{T,J}^H \in [250, 650)$ GeV and up to 20% for $p_{T,J}^H$ above 650 GeV.

Uncertainties in the modeling of the subdominant $VV$ and $\tilde{t}\tilde{t}$ backgrounds include normalization uncertainties of 80% [13] and 12% [61], respectively, as well as changes in the shape of the $m_H^j$ distribution when using different renormalization and factorization scales or alternative event generators. The effect of scale uncertainties on the $m_H^j$ distribution’s shape is also included for the $Z + \text{jets}$ background. Uncertainties in the modeling of the $Wt$ background have negligible impact.

Experimental uncertainties related to large-$R$ jets, $W/Z$ tagging, and $H \rightarrow \bar{b}b$ tagging affect the predicted event yields for the signal and the $VV$, $V + \text{jets}$, and $\tilde{t}\tilde{t}$ backgrounds. Uncertainties in the scale and resolution of the large-$R$ jet energy and mass measurements affect the shape of the $m_H^j$ distribution and are evaluated following Ref. [50].

Uncertainties in the $W/Z$ tagging are determined from studies of $\tilde{t}\tilde{t}$ events in data [57,62]. The $H \rightarrow \bar{b}b$ tagging efficiency is measured in a sample enriched in $Z \rightarrow \bar{b}b$ decays [63]. The resulting scale factors are applied to the simulation and vary between 0.86 and 1.80 across the $p_T$ bins, with uncertainties in the range 30%–60%. These scale factors are further constrained in the fit to data by the presence of the $W/Z$ resonances in the jet mass distribution.

The impact of uncertainties in the PDF sets, initial- and final-state radiation, and multiparton interactions on the signal acceptance is included. Uncertainties related to the PDF sets are derived by applying the methodology outlined by the PDF4LHC group [27] and considering four additional PDF sets (CT14, MMHT2014, NNPDF3.0, and ATLAS-epWZ12), resulting in 3% uncertainties in the signal acceptance. An uncertainty of 0.83% is applied to the integrated luminosity [64].

The signal yield and the $Z + \text{jets}$ normalization are extracted from a simultaneous binned maximum-likelihood fit to the $m_H^j$ distributions in the SR (shown in Fig. 1) and CR in the range $60$–$200$ GeV. Confidence intervals are based on the profile-likelihood-ratio test statistic [65,66]. Systematic uncertainties are implemented in the fit as nuisance parameters constrained by Gaussian or log-normal likelihood terms, and the Higgs boson mass is assumed to be $m_H = 125.09 \pm 0.24$ GeV. The overall normalization factor for the peaking component of the $Z + \text{jets}$ background (mostly from $Z \rightarrow \bar{b}b$ decays) from the fit is $1.4^{+0.8}_{-0.6}$.

The best-fit value of the signal-strength parameter, defined as the ratio of the observed signal yield to that expected in the SM, is $\mu = 1.4^{+1.0}_{-0.9}$ for the inclusive fit, corresponding to an observed (expected) significance of $1.7\sigma (1.2\sigma)$ with respect to the null signal hypothesis. An inclusive cross section of 2.24 pb [67] is used to normalize the expected signal. The statistical uncertainty in $\mu$ is 0.6, whereas the systematic uncertainty is $^{+0.8}_{-0.6}$.

FIG. 1. Higgs candidate jet mass distributions in the signal region for $p_{T,J}^H \in [250, 450)$ GeV (left), $p_{T,J}^H \in [450, 650)$ GeV (middle), and $p_{T,J}^H \geq 650$ GeV (right) obtained after the inclusive fit with a single $Z + \text{jets}$ normalization factor and a single signal strength. The bottom panels show the distributions after subtracting the multijet and top-quark backgrounds. The hatched bands show the total uncertainty in the background estimate.
The latter is dominated by the shape of the estimated multijet background and the $H \rightarrow b\bar{b}$ tagging scale factors. The corresponding inclusive cross section is $3.1 \pm 1.3 \,(\text{stat})^{+1.8}_{-1.4} \,(\text{syst}) \, \text{pb}$.

The signal strengths and cross sections are also extracted in three exclusive kinematic regions defined at generator level by $p_T^H \in [250, 450)$ GeV, $p_T^H \in [450, 650)$ GeV, and $p_T^H \geq 650$ GeV, with $|y_H| < 2$ required in each region. The results are obtained using $m_H^f$ templates extracted from the signal MC samples with the kinematic requirements applied for each region. Those results are presented in Table I.

In conclusion, a first study of associated $VH$ production is performed in the fully hadronic final state, reaching Higgs boson transverse momenta at the TeV scale. A likelihood fit to the mass distribution of Higgs candidate large-$R$ jets in events containing a $W$ or $Z$ candidate jet is used to extract the $VH$ signal both inclusively and as a function of transverse momentum. The significance of the $VH$ signal is estimated to be $1.7\sigma$ and the inclusive cross section is determined to be $3.1 \pm 1.3 \,(\text{stat})^{+1.8}_{-1.4} \,(\text{syst}) \, \text{pb}$, in agreement with the SM prediction. While the current study is limited by large uncertainties, this channel will open a kinematic region with high sensitivity to new physics contributions when larger data samples are collected.

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 have been operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPERJ, Brazil; NSERC, NRC, and CFI, Canada; CERN; ANID, Chile; CAS, MOST, and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DLR and DFNRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; NWO, Netherlands; NCN, Norway; MEiN, Poland; FCT, Portugal; MINE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF, and Cantons of Bern and Geneva, Switzerland; MOST, Taipei; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, USA. In addition, individual groups and members have received support from BCKDF, CANARIE, CRC, and DRAC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/ 013, Czech Republic; COST, ERC, ERDF, Horizon 2020, ICSC-NextGenerationEU, and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex, and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales, and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and MINERVA, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF/SFU (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), RAL (UK), and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [68].

**Appendix.**—The Monte Carlo simulation does not provide an accurate description of the QCD multijet background, especially in the high-$p_T$ kinematic region, motivating the need for a dedicated estimate of the multijet contribution in the phase space selected for this study. Therefore, this analysis exploits a fully data-driven estimation of the multijet background, using events in both the signal and control regions while testing it in validation regions. Two different methods are compared for the data-driven estimation. First, the transfer factor (TF) method uses a jet $p_T^f$ and mass-dependent transfer function to predict the yields of events that pass the event selection from the events that fail the Higgs boson tagging. The multijet background estimation and the signal extraction are performed simultaneously. Second, the boosted decision tree (BDT) method extracts the background templates from the events failing the $V$ and Higgs boson tagging. A BDT is

<table>
<thead>
<tr>
<th>Kinematic region</th>
<th>Observed $\mu$</th>
<th>Observed $\sigma$ [fb]</th>
<th>Expected $\sigma$ [fb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$250 \leq p_T^H &lt; 450$ GeV, $</td>
<td>y_H</td>
<td>&lt; 2$</td>
<td>$0.8^{+2.2}_{-1.9}$</td>
</tr>
<tr>
<td>$450 \leq p_T^H &lt; 650$ GeV, $</td>
<td>y_H</td>
<td>&lt; 2$</td>
<td>$0.4^{+1.7}_{-1.5}$</td>
</tr>
<tr>
<td>$p_T^H \geq 650$ GeV, $</td>
<td>y_H</td>
<td>&lt; 2$</td>
<td>$5.3^{+11.3}_{-3.2}$</td>
</tr>
</tbody>
</table>
used to perform a kinematic reweighting by predicting the event weights needed to bring the shapes of kinematic distributions in the control and signal regions into agreement. The following variables are used in the BDT training: Higgs candidate large-$R$ jet $p_T$, mass, pseudorapidity, and azimuthal angle, as well as the number of associated tracks and track jets.

The two methods give consistent results. After studying the fit to the sidebands and the uncertainties in the signal-plus-background Asimov fit [66] to the signal region, the
Table II. Breakdown of the various sources of uncertainty in the inclusive signal strength $\mu$.

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\delta \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal modeling</td>
<td>$+0.10$</td>
</tr>
<tr>
<td>MC statistical uncertainty</td>
<td>$-0.02$</td>
</tr>
<tr>
<td>Instrumental (pileup, luminosity)</td>
<td>$+0.13$</td>
</tr>
<tr>
<td>Large-R jet</td>
<td>$-0.13$</td>
</tr>
<tr>
<td>Top-quark modeling</td>
<td>$+0.02$</td>
</tr>
<tr>
<td>Other theory modeling</td>
<td>$0.004$</td>
</tr>
<tr>
<td>$H \rightarrow b\bar{b}$ tagging</td>
<td>$+0.12$</td>
</tr>
<tr>
<td>Multijet estimate (TF uncertainty)</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>Multijet modeling (TF vs BDT)</td>
<td>$+0.14$</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>$-0.15$</td>
</tr>
<tr>
<td>Signal statistical uncertainty</td>
<td>$-0.05$</td>
</tr>
<tr>
<td>$Z +$ jets normalization</td>
<td>$-0.03$</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>$-0.03$</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>$+1.02$</td>
</tr>
</tbody>
</table>

The impact of the different sources of uncertainty on the signal strength is presented in Table II.

[17] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector and the $z$ axis along the beam pipe. The $x$ axis points to the center of the LHC ring, and the $y$ axis points upward. Cylindrical coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the $z$ axis. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.


