Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle $X$ in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector

Aad, G.; ATLAS Collaboration

DOI
10.1103/PhysRevD.108.052009

Publication date
2023

Document Version
Final published version

Published in
Physical Review D. Particles and Fields

License
CC BY

Citation for published version (APA):
Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle $X$ in hadronic final states using $\sqrt{s} = 13$ TeV $pp$ collisions with the ATLAS detector

G. Aad et al.*
(ATLAS Collaboration)

(Received 12 June 2023; accepted 29 July 2023; published 18 September 2023)

A search is presented for a heavy resonance $Y$ decaying into a Standard Model Higgs boson $H$ and a new particle $X$ in a fully hadronic final state. The full Large Hadron Collider run 2 dataset of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS detector from 2015 to 2018 is used and corresponds to an integrated luminosity of 139 fb$^{-1}$. The search targets the high $Y$-mass region, where the $H$ and $X$ have a significant Lorentz boost in the laboratory frame. A novel application of anomaly detection is used to define a general signal region, where events are selected solely because of their incompatibility with a learned background-only model. It is constructed using a jet-level tagger for signal-model-independent selection of the boosted $X$ particle, representing the first application of fully unsupervised machine learning to an ATLAS analysis. Two additional signal regions are implemented to target a benchmark $X$ decay into two quarks, covering topologies where the $X$ is reconstructed as either a single large-radius jet or two small-radius jets. The analysis selects Higgs boson decays into $b\bar{b}$, and a dedicated neural-network-based tagger provides sensitivity to the boosted heavy-flavor topology. No significant excess of data over the expected background is observed, and the results are presented as upper limits on the production cross section $\sigma(pp \rightarrow Y \rightarrow XH \rightarrow q\bar{q}bb)$ for signals with $m_Y$ between 1.5 and 6 TeV and $m_X$ between 65 and 3000 GeV.

DOI: 10.1103/PhysRevD.108.052009

I. INTRODUCTION

The Standard Model (SM) provides a framework for understanding fundamental particles and interactions that has been remarkably predictive of experimental results over several decades. The discovery of the Higgs boson in 2012 [1,2] completed the set of particles predicted by the SM.

The sensitivity of the Higgs boson mass to radiative corrections implies either extreme fine-tuning in the SM or the existence of new particles at an energy scale not far above the Higgs boson mass. This theoretical motivation, coupled with the existing experimental mass reach of the Large Hadron Collider (LHC) at CERN, motivates searches for new particles with $\mathcal{O}(\text{TeV})$ masses. Because the Higgs boson couples more strongly to heavier particles, it is natural to expect that these new heavy particles may have decays to a Higgs boson.

A search is presented here for a new TeV-scale narrow-width particle $Y$, which decays into a Standard Model Higgs boson $H$ and a new particle $X$ with a mass near the weak scale. A fully hadronic final state is targeted for both particles. Tagging of the boosted Higgs decay into two $b$ quarks ($H \rightarrow b\bar{b}$ tagging) enhances a signal by using the highest branching fraction decay of the Higgs boson. A novel jet-level implementation of anomaly detection implemented via an unsupervised machine-learning architecture is used to select $X$ particles solely because of their incompatibility with the expected SM background.

The application of anomaly detection to collider searches is a rapidly growing effort in the high-energy physics community (see Refs. [3–5] for overviews). The lack of evidence for new interactions and particles since the Higgs boson’s discovery has motivated the execution of generic searches to complement the existing rigorous, model-dependent analysis program. Machine learning provides an excellent framework for the construction of tools that can isolate events in data solely because of their incompatibility with a background-only hypothesis. Building a tool to perform model-independent classification of collision events involves training on data events and, therefore, requires the ability to cope with a lack of labels indicating whether inputs are signal or background.
symmetry constraints with an isospin SU(2) triplet formed of a neutral $Z'$ and two charged $W^{\pm}$ bosons.

This search uses the full LHC run 2 $\sqrt{s} = 13$ TeV $pp$ dataset collected by the ATLAS detector from 2015 to 2018 and corresponding to an integrated luminosity of 139 fb$^{-1}$. A search for the $Y \rightarrow XH$ process was previously performed by ATLAS using 36.1 fb$^{-1}$ of data, assuming $X \rightarrow q\bar{q}$, and found no significant excess for $Y$ masses from 1 to 4 TeV and $X$ masses from 50 to 1000 GeV [12]. In addition to the much larger dataset, the present analysis includes several key improvements relative to the previous search, such as a neural-network-based tagger optimized for the boosted $H \rightarrow b\bar{b}$ topology, anomaly detection for enhanced signal model independence, and the use of two orthogonal regions to capture both boosted and resolved reconstruction of the nominal $X$ decay into two quarks. The ATLAS Collaboration previously leveraged anomaly detection implemented with weakly supervised machine learning in a search for narrow new resonances in dijet events using the full run 2 dataset [13]. The CMS experiment performed a multidimensional search for diboson resonances with their full run 2 dataset of 138 fb$^{-1}$ which includes sensitivity to the signature discussed here [14]. A maximum local (global) significance of 3.6 (2.3) standard deviations was observed for two mild excesses of events, at resonance masses of 2.1 and 2.9 TeV with $X$ masses consistent with $W$, $Z$, or Higgs bosons.

The analysis regions in the present search are built by selecting two large-$R$ jets, with additional criteria to reduce the background to events with Higgs and $X$ particles in the search regions. Three overlapping analysis categories are defined with different $X$ selections, namely the anomaly region and two orthogonal regions optimized for the benchmark $X \rightarrow q\bar{q}$ decay. The degree of model independence of the $X$ tagging is assessed using three simulated signals with different decays of the $X$ particle that lead to differing large-$R$-jet topologies, in addition to the two-light-quark final state used to generate the $(m_{\gamma}, m_{X})$ signal grid. The background to the signal process consists primarily of multijet events from quantum chromodynamics (QCD) processes. It is estimated with a fully data-driven method that incorporates reweighting based on a deep neural network (DNN) to ensure good modeling. The final fits are made to the reconstructed invariant mass distribution of the $Y$ candidates in overlapping ranges of $X$ candidate mass to increase the sensitivity to a signal. Since no significant deviation from the estimated background is observed, results are presented as upper limits on the cross section times branching fraction for the generic HVT process.

II. ATLAS DETECTOR

The ATLAS detector [15] is a multipurpose particle detector with a forward-backward symmetric cylindrical
geometry and a near $4\pi$ coverage in solid angle. The inner-detector (ID) system is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four space-point measurements per track, the first hit normally being in the insertable B layer installed before run 2 [16,17]. The next detector outward is the silicon microstrip tracker, which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker, which enables radially extended track reconstruction up to $|\eta| = 2.0$. Lead and liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A steel and scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end cap and forward regions are instrumented with LAr calorimeters for both the EM and hadronic energy measurements up to $|\eta| < 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. The muon spectrometer includes a system of precision chambers for tracking and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100 kHz on average depending on the data-taking conditions. This is followed by a software-based high-level trigger that reduces the rate of selected events to 1 kHz for offline storage. An extensive software suite [18] 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.

III. DATA AND MONTE CARLO SIMULATION

This search is performed with $\sqrt{s} = 13$ TeV $pp$ collision data collected by the ATLAS detector during run 2 of the LHC (2015–2018). After the application of data quality requirements [19] to ensure that all detector components are operating normally, the dataset corresponds to an integrated luminosity of $139.0 \pm 2.4$ fb$^{-1}$ [20,21]. Data events used in this analysis were triggered by the presence of a high-$p_T$ large-$R$ jet, built at trigger level using calorimeter-cell energy clusters calibrated to the hadronic scale utilizing the local cell-signal weighting method [22]. The unprescaled single-large-$R$-jet trigger with the lowest jet-$p_T$ threshold was used in each year, corresponding to thresholds of 360 GeV in 2015, 420 GeV in 2016, and 440 GeV in 2017 and 2018. Further offline selection criteria are imposed (see Sec. VI) to ensure that selected events are all in the kinematic region where the trigger is fully efficient.

Monte Carlo (MC) event generators were used to simulate the signal targeted by this search. The simulated event samples include the effect of additional $pp$ interactions (pile-up) in the same or neighboring bunch crossings. The effect is assessed by overlaying simulated minimum-bias events on each hard-scatter event and accounting for how interactions in the previous or following bunch crossing affect the detector response. The simulated events are weighted to reproduce the pileup distribution observed in each data-taking period. The detector response was simulated [23] using Geant4 [24] and the events were reconstructed with the same software as used for data.

A generic HVT [11,25] model with simulated $qq$ scattering is used as a baseline for the $Y \to XH$ signal samples. The $Y$ resonance is assumed to have a width that is narrow compared to the detector resolution. The generation of the $W' \to WH$ process was modified to replace the $W$ boson with a new spin-1 boson $X$ with a width of 2 GeV and a 100% branching fraction to $u\bar{d}$. The Standard Model Higgs boson was generated with decays into $b\bar{b}$ only. Samples with $Y$ resonance masses of 1–6 TeV were generated using MadGraph5_aMC@NLO 2.7.2 and 2.7.3 [26] interfaced to PYTHIA 8.2 [27] for parton showering and hadronization with the NNPDF2.3LO parton distribution function (PDF) set [28], at leading order in QCD, and parameter values set according to the A14 tune [29]. A two-dimensional grid was generated for 195 signal points defined by $Y$ and $X$ mass values, with $Y$ masses in the range 1.5–6 TeV and $X$ masses in the range 65–3000 GeV. The $Y$ and $X$ mass values account for the kinematic constraint $m_Y > m_X$, and the $X$ mass boundaries are dictated by achievable sensitivity. The signal acceptance times the efficiency has values ranging from a few percent to 40% for the various signal region selections.

In addition to the $Y \to XH$ signal grid, three signals with different jet topologies were used to craft and assess the degree of model independence in the $X$ anomaly-tagging procedure. They were all generated with PYTHIA 8 using the same tune (A14) and PDFs (NNPDF2.3LO) as the $Y \to XH$ signal grid. The HVT $W' \to WZ$ configuration from Ref. [30] was used to create generic resonance signatures of the form $A \to BC$, where the daughter particles decay into three light quarks or two heavy-flavor quarks. Choosing as benchmarks a parent $A$ mass of 3 TeV and daughter masses of 200 and 400 GeV creates approximately the same kinematic scenario as the $Y \to XH$ phase space of interest, leading to boosted decay products that are captured as large-$R$ jets. Depending on the daughter decay

---

1ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the $z$ axis along the beam pipe. The $x$ axis points from the IP to the center of the LHC ring, and $y$ 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 $\eta = -\ln \tan(\theta/2)$. 

---

PHYS. REV. D 108, 052009 (2023)
into either $qqq$ or $b\bar{b}$, the large-$R$ jet is produced with three-prong substructure or displaced vertices, respectively. Lastly, the PYTHIA 8 hidden valley model $A$ configuration was used to generate a “dark jet” signal that arises from the decay of a $Z'$ dark sector mediator into two dark quarks, where the $Z'$ has the same benchmark mass of 3 TeV [31]. These jets contain dark matter particles that do not interact with the detector, creating a hadronization pattern that contains both visible and invisible energy. Such jets are difficult to recognize using traditional substructure variables that focus on specific signal topologies, making them an ideal target for anomaly detection based on background-only characterization.

### IV. OBJECT SELECTION

Charged-particle tracks measured to have $p_T > 500$ MeV in the inner detector are used to reconstruct $pp$ collision vertices, and the one with the largest $\Sigma p_T^2$ of such tracks is chosen as the event’s primary vertex.

Jets are built with the anti-$k_t$ algorithm [32] in FastJet [33], using two radius parameter values to cluster the constituent detector signals: $R = 1.0$ for large-$R$ jets $J$ and $R = 0.4$ for small-$R$ jets $j$. Large-$R$ jets must satisfy $p_T > 200$ GeV and $|\eta| < 2.0$. They are constructed from combinations of tracks and calibrated calorimeter energy clusters known as track-caloclusters (TCCs) [34]. The TCCs take advantage of both the excellent energy resolution of the ATLAS calorimeter and the better angular resolution of the ID tracking system at very high energy, where the calorimeter is less able to discern jet substructure. This is particularly advantageous for measuring substructure in highly boosted jets, as is needed in this search to distinguish $Y \to XH$ signal from multijet background. To minimize the impact of pile-up, large-$R$ jets are trimmed [35] by removing any $R = 0.2$ subjects carrying less than 5% of the large-$R$ jet’s $p_T$. The jet energy and mass is calibrated with a MC-based method [36].

Small-$R$ jets are built from particle-flow objects [37], with improved accuracy of the jet’s charged-hadron energy component through measurements of the associated charged-particle momenta in the ID. Small-$R$ jets must have $p_T > 20$ GeV and $|\eta| < 4.5$. An additional event-level veto is applied to reject events with a jet that is likely to have come from calorimeter noise, beam-induced background, or a cosmic ray [38].

Leptons, while excluded from the analysis region selections, participate in an overlap removal procedure that prevents double counting of overlapping objects. Leptons must have $p_T > 7$ GeV, with $|\eta| < 2.47$ for electrons and $|\eta| < 2.7$ for muons. To ensure that each lepton’s track originates from the primary vertex, the transverse impact parameter $d_0$, measured relative to the beam line, must have a significance $|d_0/\sigma(d_0)| < 3(5)$ for muons (electrons), and the longitudinal impact parameter $z_0$, from the primary vertex to the point where $d_0$ is measured, must satisfy $|z_0 \sin \theta| < 0.5$ mm. No lepton isolation criteria are applied.

The overlap removal procedure is applied to all selected leptons and jets. If two electrons share the same track, or the separation between their two energy clusters satisfies $|\Delta \eta| < 0.075$ and $|\Delta \phi| < 0.125$, then the lower-$p_T$ electron is discarded. Electrons that fall within $\Delta R = 0.02$ of a selected muon are also discarded. For electrons and nearby small-$R$ jets, the jet is removed if the separation between the electron and jet satisfies $\Delta R < 0.2$; the electron is removed if the separation satisfies $0.2 < \Delta R < 0.4$. For muons and nearby small-$R$ jets, the jet is removed if the separation between the muon and jet satisfies $\Delta R < 0.2$ and if the jet has less than three tracks or the energy and momentum differences between the muon and the jet are small; otherwise the muon is removed if the separation satisfies $\Delta R < 0.4$. To prevent double counting of energy from an electron inside a large-$R$ jet, the large-$R$ jet is removed if its separation from the electron satisfies $\Delta R < 1.0$.

### V. VARIATIONAL RECURRENT NEURAL NETWORK FOR ANOMALOUS-JET TAGGING

The anomaly detection in this search is performed with a jet-level anomaly score ($S_\Delta$) [39]. The $S_\Delta$ value is given by a variational recurrent neural network (VRNN) [40], which consists of a variational autoencoder (VAE) whose latent space is updated at each time step of a recurrent neural network (RNN). The key features of this tool as it pertains to this analysis application are provided below, with a comprehensive description provided in Ref. [39].

The class of models that are based on autoencoders are popular for the selection of rare or outlier events in high-energy physics [41–46]. Such an architecture is composed of two separate stages: an encoder that compresses the input into a lower-dimensional latent space through identification of its salient features and a decoder that samples from the latent space and attempts to reproduce the input in its original dimensionality. Minimizing the reconstruction error between input and output is a key element of the training and allows the autoencoder to implicitly learn the underlying distribution of features in its input dataset and, thus, recognize anomalous data through its inability to reconstruct such events with low error.

VAEs extend this concept to perform Bayesian inference via a latent space that can accommodate a distribution of encoded information. The VRNN architecture combines the variational inference capabilities of a VAE with the sequence modeling provided by an RNN. The inclusion of a learned, time-dependent prior distribution is an essential feature of the VRNN and allows the modeling of complex structured sequences with high variability.

The VRNN is trained on large-$R$ TCC jets in the ATLAS Run 2 dataset satisfying the criteria in Sec. VI for full trigger efficiency, the jet requirements described in Sec. IV, and $p_T(J) > 1.2$ TeV. Its architecture is the same as
that described in Ref. [39]. Training is performed using the PyTorch deep-learning library [47], and the network is updated using the Adam optimizer [48] with a learning rate parameter of $10^{-5}$. No regularization via weight decay is applied, but gradient clipping is used with a clip value of 10.

The input jets are modeled as a sequence of up to 20 constituent four-vectors per jet, ordered in $k_t$ splitting starting from the highest-$p_T$ constituent. The $p_T$ selection is designed to chose input jets with highly boosted topologies, for which the VRNN is expected to most improve the sensitivity over the substructure-based approach, as these topologies are well described by $k_t$-sorted sequence modeling and are also challenging to distinguish with substructure methods. The training is conditioned on four high-level variables, namely the energy-correlator substructure variable $D_2$ [49,50] for two-prong sensitivity, the $N$-subjettiness ratio $\tau_2$ [51] for three-prong sensitivity, and the two $k_t$-splitting-scale ratios $d_{12}$ and $d_{23}$ [52]. An alignment procedure applied to each jet rescales to the same $p_T$, boosts to the same energy, and rotates to the same orientation in $\eta$ and $\phi$, inhibiting the VRNN from selecting jets with rare kinematic properties without considering internal constituent structure. This input modeling is designed to reveal correlations between constituents and substructure, allowing the VRNN to distinguish jets with anomalous energy-deposition patterns from the background of homogenous jets originating from QCD processes.

Since the input consists solely of jets from data, no labeling scheme is used in training, distinguishing this method of unsupervised learning from traditional supervised machine learning, where the input is labeled in signal or background categories. Of the data meeting the above criteria, 90% is used for training and 10% for validation. No blinding criterion is applied to the training data, so events or background categories. Of the data meeting the above criteria, 90% is used for training and 10% for validation. No blinding criterion is applied to the training data, so events that are used in training may also enter the final analysis regions. Since the VRNN performance is consistent across a variety of signal contamination fractions in training [39], and the presence of signal in the training dataset would only lower the sensitivity and cannot produce a false excess, biasing the final result is not a concern.

The loss function of the VRNN is composed of two terms: a reconstruction error term, to minimize differences between the decoded result and the original input, and the Kullback-Leibler (KL) divergence $D_{KL}$ [53] of the encoded approximate posterior distribution from the Gaussian prior distribution. The loss $\mathcal{L}(t)$ for each time step $t$ is given by Eq. (1), where $\mathbf{x}(t)$ is the input constituent four-vector, $\mathbf{y}(t)$ is the output constituent four-vector, $z_t$ is the approximate posterior, $z_t$ is the learned prior, and $\lambda$ is a constant parameter to weight the $D_{KL}$ contribution:

$$\mathcal{L}(t) = |\mathbf{y}(t) - \mathbf{x}(t)|^2 + \lambda D_{KL}(z_t||z_t).$$

Here, the time step of the RNN refers to the location in the input sequence. The learned prior $z_t$ is a function of the current time step’s hidden state, and the hidden state is updated via a recurrence relation after each time step. An overall loss $\mathcal{L}$ over the sequence is then computed by averaging the individual time-step losses over the length of the sequence.

The KL divergence of the encoded posterior distribution from the true posterior distribution is jointly minimized along with the loss $\mathcal{L}(t)$, ensuring that the architecture accurately describes the underlying distribution of data [54]. The $D_{KL}$ term has been shown to provide better discrimination between anomalous and standard jets than either the reconstruction error or the loss term as a whole. The anomaly score is therefore defined in terms of the $D_{KL}$ of each constituent, averaged over the whole jet, and restricted to the range of (0,1) via exponentiation, as shown in Eq. (2). The resulting variable $S_A$ ensures that more-anomalous jets populate higher values and background populates lower values and is defined as

$$S_A = 1 - e^{-D_{KL}}.$$  

The performance of the VRNN can be assessed by applying Eq. (2) to both background and a variety of signal models and studying the discrimination power of $S_A$. Its performance in this analysis and the use of the anomaly score to select the $X$ jet are described in the following section.

VI. EVENT SELECTION

The experimental signature of the $Y \rightarrow XH$ signal contains at least two jets with high transverse momentum. Signal-like events are selected through event-level criteria, along with jet-tagging criteria for both the $X$ and $H$ boson. In total, the analysis is performed three times, once for each of the three signal regions (SRs). The three analysis categories are defined using the selection criteria for the $X$ particle. Each signal region utilizes five background estimation regions, composed of three control regions (CRs) and two validation regions (VRs), which are defined using common selections on the Higgs boson. The definitions of these regions and motivation for the selection criteria are provided in this section.

The analysis uses data events passing a high-$p_T$-threshold single large-$R$-jet trigger. Selected events must have a primary vertex and at least two large-$R$ jets. Only the two highest-$p_T$ large-$R$ jets in an event are considered further, and at least one of their masses must exceed 50 GeV to remove poorly reconstructed events. To ensure that the trigger is fully efficient, the invariant mass of the two large-$R$ jets must satisfy $m_{jj} > 1.3$ TeV and the leading large-$R$ jet, $J_1$, must have $p_T(J_1) > 500$ GeV. These requirements, along with a selection on the $D_{H_{mb}}$ score defined in the next subsection, compose the analysis preselection.

052009-5
As mentioned previously, the three analysis SRs vary in the selection criteria for the $X$ particle. The primary SR, using the VRNN for generic $X$ tagging, is referred to as the anomaly SR. The other two SRs target the benchmark $X \rightarrow q\bar{q}$ decay and are thus called the two-prong SRs. Here the jet substructure is used to distinguish reconstruction of the $X$ particle as either a single large-$R$ jet (merged SR) or two small-$R$ jets (resolved SR). They are not required to be orthogonal to the anomaly SR but do not overlap with each other. The analysis CRs include the high sideband (HSB) of the Higgs boson candidate mass $m_H$, used to generate the DNN reweighting for the background estimation procedure, and CR0, which provides the final background template for the SR. The VRs are defined in the low sideband (LSB) of the Higgs boson candidate mass and validate the reweighting derived from the HSBs. The full analysis flow is shown in Fig. 2, with selection details provided below.

**A. X/H ambiguity resolution**

Since the SRs are constructed by placing requirements on the different properties of the $X$ and $H$ large-$R$ jets, an ambiguity must be resolved to determine which of the two $J$'s in the event is more likely to be the Higgs boson and thus subject to the Higgs boson selection criteria. This is done by using a classifier based on a neural network (NN) to separate bosons decaying into $b\bar{b}$ from top-quark and QCD jets [55]. The tagger is trained on the large-$R$ jet $p_T$ and $\eta$, along with the subjett flavor-tagging score $DL_{1r}$ [56] for up to three track subjets constructed with a jet-$p_T$-dependent radius parameter [57]. The tagger version used here includes a reweighting of all training inputs to have the same $p_T$ and $\eta$ distributions, to minimize biasing of the tagger toward high-$p_T$ or central jets, respectively. The outputs of the NN are three classification scores corresponding to the probability for the jet to have originated from a Higgs boson ($P_{Higgs}$), top quark ($P_{top}$), or multijet process ($P_{multijet}$), and these are subsequently combined into the jet-level discriminant $D_{Hbb}$:

$$D_{Hbb} = \ln \left( \frac{P_{Higgs}}{f_{top} \cdot P_{top} + (1 - f_{top}) \cdot P_{multijet}} \right).$$

In Eq. (3), $f_{top}$ determines the weight assigned to the top-quark background shape in the final discriminant, and its chosen value of 0.25 is based on signal-to-noise optimization studies.

To perform the large-$R$ jet ambiguity resolution, $D_{Hbb}$ is computed for both $J$’s in the event. The jet with the larger value of $D_{Hbb}$ is labeled as the Higgs boson candidate ($J_H$), and the other $J$ is by default the $X$ candidate ($J_X$), thus determining which jet is subject to further $H$ and $X$ selection. This procedure has an accuracy of over 90% in the highly boosted region of the signal grid.
Across the full pT range, using the anomaly score value for the two-prong resolved regions. The anomaly region is defined using lous-jet content in the anomaly, two-prong merged, and dependent selection is applied to enhance the anomaly. Requiring 60% efficient for the selection of the boosted particles. Requiring 60% efficient for the selection of the boosted mass points where the boosted \( H \rightarrow b \bar{b} \) topology across the full pT range. All distributions are normalized to unity.

\( (m_X/m_Y < 0.5) \), decreasing to no less than 75% for less boosted mass points where the boosted \( H \rightarrow b \bar{b} \) tagging is less efficient. The resulting distribution of \( D_{H\beta} \) for the \( J \) chosen as the Higgs boson candidate, for both data and representative \( Y \rightarrow XH \) signals at preselection, is shown in Fig. 3. Additionally, a preselection of \( D_{H\beta} > -2 \) is applied to remove events that are determined to be not \( H \rightarrow b \bar{b} \)-like, thus ensuring the data-driven background estimation focuses on a phase-space region that is close to that of the signal.

**B. X selection**

After the \( X \) candidate’s \( J \) is determined, a region-dependent selection is applied to enhance the anomalous-jet content in the anomaly, two-prong merged, and two-prong resolved regions. The anomaly region is defined using the anomaly score value \( S_A \) of the \( X \) jet as determined by the VRNN. Figure 4 shows the resulting distribution of \( X \) candidate \( S_A \) values after the analysis preselection is applied, comparing data with three different \( Y \rightarrow XH \) baseline signal samples as well as the three signals with alternative \( X \) decays. The absence of signal information in the construction of the score means that it is not expected to outperform analytic or supervised methods for any single signal model. However, using the VRNN approach to characterize the background distribution provides broad sensitivity to a variety of signals that differ from jets originating from QCD processes. The most notable difference in \( S_A \) shape is found for dark jets, which have a substructure that is not well characterized by existing variables. The dependence of the sensitivity on kinematics is also relevant to performance, as the \( S_A \) distribution for the highly resolved point \( (m_Y = 5000 \text{GeV}, m_X = 2500 \text{GeV}) \) populates lower, and thus less anomalous, values than the data, making this signal indistinguishable from background in the anomaly detection approach. A selection \( S_A > 0.5 \) is chosen for the \( J_X \) candidate and provides modest but comparable discrimination power in data for the disparate signal-jet topologies considered in the full expected pT range of the \( X \) candidate jets. This enhances the signal-to-background ratio for the two boosted points, \( (m_Y = 2000 \text{GeV}, m_X = 300 \text{GeV}) \) and \( (m_Y = 3400 \text{GeV}, m_X = 110 \text{GeV}) \), by approximately 25% relative to the inclusive selection.

Regions focusing on the benchmark \( X \rightarrow q\bar{q} \) decay are used to supplement the anomaly detection analysis and interpret the results. The merged region is defined by applying a selection \( D_{2\text{trk}} < 1.2 \), where \( D_{2\text{trk}} \) is the same as \( D_2 \) but computed using only tracks associated with the jet. The choice of using only tracks in the \( D_2 \) calculation is motivated by the ability to propagate track-only uncertainties into the final signal efficiency. Detailed comparisons of \( D_{2\text{trk}} \) with the standard \( D_2 \) variable at analysis preselection have shown \( D_{2\text{trk}} \) and \( D_2 \) to be highly compatible, ensuring its sensitivity to two-prong signals with respect to multijet background. Because jets with
two-prong substructure have lower values of $D_{2k}$, this selection enhances the presence of fully merged decays where the $X$ decay products are well contained by a single $J$. Since the efficiency of substructure variables is mainly driven by the boost of the large-$R$ jet and therefore its $p_T$, the optimal cut value of 1.2 was chosen so as to maximize signal-versus-background discrimination across the $p_T$ range of the $X$ candidate, combining all mass points in the $Y \to XH$ grid. Distributions of $D_{2k}$ for $J_X$ are shown in Fig. 5 for both data and representative $Y \to XH$ signals at preselection.

Due to kinematic constraints driven by the $m_X/m_Y$ ratio, events that fail the merged selection typically also show poor reconstruction of the $X$ mass by the large-$R$ jet. In these events, the large-$R$ jet does not sufficiently contain all of the $X$ decay products, leading to an underestimation of the $X$ particle mass. To achieve better sensitivity in this regime, a resolved selection is defined by requiring the $D_{2k}$ value of $J_X$ to be greater than 1.2, corresponding to a less boosted $X$ particle which is more appropriately reconstructed using two small-$R$ jets. A dedicated algorithm matches two small-$R$ jets to the $X$ by requiring at least four $J$ in the event, discarding the two with the smallest $\Delta R$ to the Higgs boson candidate $J_H$, and selecting the two remaining small-$R$ jets with the highest $p_T$. The mass of the $Y$ resonance is then computed using the Higgs boson candidate large-$R$ jet and the two small-$R$ jets matched to the $X$. Additional requirements of $|\Delta y_{J, J}| < 2.5$ and a $p_T$ imbalance $p_T^{\text{bal}} < 0.8$ are applied to aid accurate resolved $X$ reconstruction; here $\Delta y$ refers to separation in rapidity and $p_T^{\text{bal}}$ is defined as the difference of the $X$ particle’s small-$R$ jet transverse momenta divided by their sum. The resolved region improves the signal efficiency for points with higher $m_X/m_Y$ (corresponding to less boost for the $X$ candidate decay) by a factor of approximately 5 relative to the boosted selection.

### C. H selection

The Higgs boson tagging is performed after the $J_X$ selection and sorts the events into the three analysis categories, for which the background is estimated. For all signal regions, a requirement $D_{H_{bb}} > 2.44$ providing a constant 60% efficiency across jet $p_T$ is applied to the Higgs boson candidate $J_H$, along with a mass window requirement of $75 < m_H < 145$ GeV. The data-driven background estimation uses events from the HSB of the Higgs boson candidate mass ($145 < m_H < 200$ GeV), which are further separated into HSB0 and HSB1 depending on whether the Higgs boson candidate $J_H$ fails or passes the $D_{H_{bb}}$ selection. Validation is performed in the LSB, where the reconstructed Higgs boson mass must be between 65 and 75 GeV. LSB0 and LSB1 are similarly defined as having Higgs boson candidates that fail or pass the $D_{H_{bb}}$ tagging criterion, respectively. CR0 is defined as the set of events where the $H$ boson is in the SR mass window but fails the $D_{H_{bb}}$ tagging. DNN-based reweighting is applied to CR0 to obtain the final background prediction in the SR. The contamination from signal events is found to be of the order of a few percent in the HSB and no more than 10% in the LSB, for all mass points in the $Y \to XH$ grid.

In total, three SRs and 15 background estimation regions, five for each SR, are used in the analysis. A summary of all region definitions is provided in Table I.

### D. Binning optimization

The signal search in the two-dimensional space of $m_Y$ versus $m_X$ applies a variable-width sliding window to the $J_X$ candidate mass spectrum, dividing the data into a series of overlapping $m_X$ ranges in which the $m_{jj}$ distribution is fitted. The chosen binning for $m_Y$ and $m_X$ is based on the ATLAS detector’s mass resolution for a generic narrow-width resonance, given by the fitted width of the Gaussian core of a double-sided Crystal Ball function. Modifications are made to the binning to account for the limited number of data events. In the $m_Y$ spectrum, bins are widened at higher values of $m_Y$ to ensure that at least one event is present in each bin when using LSB1 data. For the $m_X$ categories, an initial set of mass windows is generated from a linear fit of $m_X$ resolution as a function of $m_Y$. The first window is chosen to have a bin center at $m_X = 65$ GeV, the lowest generated signal $X$ mass, and its width is chosen to be twice the resolution obtained from the fit. Subsequent windows, increasing in $m_X$, are chosen in the same way, where their bin centers are selected to be higher than the previous window’s bin center by an amount that is equivalent to half of the previous window’s mass resolution. The widths of high-mass windows are expanded...
The reweighting function is defined as the ratio of the multidimensional probability distribution function of data in HSB1 to that of data in HSB0. In this analysis, the statistical procedure of *direct importance estimation* [58,59] is utilized, where the ratio is estimated directly from data without having to analytically compute each individual PDF. It is implemented via the training of a DNN, where a custom square-root loss function [60] is chosen to produce weights that can accurately reproduce the observed ratio in data. The DNN outputs event-level weights that are assumed to be approximately independent of $m_H$, an assumption that can be verified using events in the LSB, and thus can be applied to an untagged region to produce the $m_{jj}$ shape in the corresponding $H \rightarrow b\bar{b}$-tagged region.

Events are considered for training the DNN if they pass the analysis preselection, satisfy $145 < m_H < 175$ GeV, and additionally have at least two track jets associated with the Higgs boson candidate $J$. They are modeled as a set of variables, namely the transverse momentum, pseudorapidity, azimuthal angle $\phi$ and energy of the Higgs boson candidate; the number of tracks associated with the Higgs boson candidate; and the transverse momentum, pseudorapidity, $\phi$ and mass of the first two track jets associated with the Higgs boson candidate, ordered in $p_T$. Each variable $x$ is standardized with the transformation $x = (x - \mu)/\sigma$, where $\mu$ and $\sigma$ are the mean and the standard deviation of the distribution of the $x$ variable.

The DNN is built using a fully connected sequential model from *Keras* [61] with three inner layers, each with 20 neurons and a rectified linear unit activation function. In order to reduce the problem of overfitting during training, 10% of connections among inner layers are randomly severed (called “dropout”). The last layer has a single output with a simple linear activation function. The model is trained using the Adam optimizer [48] in *Keras*, with *TensorFlow* [62] as its back end. Training is performed using a batch size equal to the full dataset size for 1600 epochs, with early stopping at 100 epochs if the value of the loss calculated on the validation dataset does not decrease for 100 subsequent epochs.

The output weights of the DNN are validated using data in the LSB. Figure 6 shows the impact of the reweighting on distributions of several key analysis variables, using the two-prong merged LSB VR as an example. Three curves are shown for each variable, comparing the LSB0 data, before and after DNN reweighting is applied, with the target data distribution in LSB1. These variables are chosen so as to focus on kinematic variables of the Higgs boson, whose properties are used to distinguish the CR from the SR, and variables involved in the final statistical treatment. Good agreement between the reweighted shape and the true tagged data is observed for each variable, suggesting a robust background model. Since the training is performed...
inclusively in the X tagging, the same conclusion holds for the anomaly and two-prong resolved LSB regions, which is verified in comparisons of distributions between the reweighted LSB0 and LSB1 in those regions. Also, since the reweighting is applied to \( m_{jj} \) distributions after the X-tagging selection is applied, no extrapolation across \( D_{trk}^2 \) or \( S_A \) is required, and the method shows similar background modeling across all three SRs. Any minor residual differences between predicted background and data in the SRs are covered by the nonclosure systematic uncertainty, described in Sec. VIII.

VIII. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties affect both the data-driven background estimate and the simulated signal. All background uncertainties are estimated using an \( m_{jj} \) shape that is inclusive in \( m_X \) and applied to each \( m_X \) category fully correlated among \( m_{jj} \) bins. Background shape variations are derived from three sources.

The first is from training the DNN in an alternative region of \( 165 < m_H < 200 \) GeV to account for the way in which phase-space uncertainties may affect the obtained weights. This region has approximately the same number of events and tagging efficiency as the nominal training region, helping to isolate the way in which changing only the DNN model affects the weights. Upward and downward variations are defined by symmetrizing the shape difference in \( m_{jj} \) between the two different models, yielding an effect of \( O(1-10\%) \) across the distribution.

Another DNN variation accounts for the finite size of the training sample and the random initialization of the weights. It is estimated with a bootstrap procedure [63,64] where a set of 100 bootstrap networks are trained. For each training, the training dataset is varied by resampling it with replacement, assigning to each event a new weight.
that comes from a Poisson variable with a mean of one. Two additional templates are formed to create a symmetric error band by taking the median weight for each event and adding or subtracting half of the interquartile range. This corresponds to a $O(1)\%$ effect across $m_y$.

Lastly, a nonclosure uncertainty is included to cover model discrepancies that may arise when weights derived from the NN training in the HSB are extrapolated to the LSB and subsequently to the SR. It is defined by the symmetrized shape difference between the data and predicted background in the LSB. To avoid being sensitive to statistical fluctuations in the LSB, smoothing is applied to the variation, where it is rebinned to reduce the relative statistical uncertainty. The nonclosure uncertainty is negligible for low $m_{jj}$ and rises to $O(10)\%$ in the $m_{jj}$ tails.

Normalization and shape uncertainties are applied to the simulated signals. An uncertainty of 1.7% is applied to the luminosity measured with the LUCID-2 detector\cite{65}. The uncertainty from the trigger selection is negligible, and thus not included, since the requirement of $m_{jj} > 1.3$ TeV ensures that the trigger is fully efficient.

Scale factors (SFs) computed to match the $H \to b\bar{b}$ tagger efficiency between data and simulation are applied to the signal template, and their uncertainties are similarly propagated to the signal normalization. These SFs are computed using the methodology described in Ref.\cite{66}, with the substitution of an updated $D_{\text{hts}}$ that includes the $\eta$ reweighting of inputs as described in Sec. VI. They are binned in large-$R$-jet $p_T$, where the SF for the highest $p_T$ bin is extrapolated to cover the upper end of the $p_T$ region probed by the analysis selection. Agreement between data and MC simulation for these SFs is verified with $Z \to b\bar{b}$ events in calibration studies. The SFs are bounded by approximately 1.1 and 1.4 across all $p_T$ bins, with uncertainties in the range of approximately ±(0.3–0.5).

Instrumental systematic uncertainties arise from uncertainty in the jet $p_T$ scale and resolution, for both the large-$R$ and small-$R$ jets, and affect the shape of the $m_{jj}$ distribution. Uncertainty in the large-$R$-jet $p_T$ scale is an important effect in the search for resonant structures in the presence of rapidly falling background spectra, as it can shift the peak of the resonance. It is evaluated using tracking-to-calorimetric $p_T$ double ratios between data and simulation, where any observed differences are assigned as baseline systematic uncertainties\cite{67}. Even though the analysis relies on jets built from TCCs, the total $p_T$ of the jet is still solely derived from calorimeter information, keeping it independent of the track-based jet $p_T$. Past analyses have studied the possible impact of calorimeter versus track-based $p_T$ by cross-calibrating between per-jet TCC and calorimeter $p_T$ and found it to be negligible\cite{30}. The total jet-$p_T$-scale uncertainty varies with $p_T$ and $\eta$ and is typically around ±5%. Uncertainties from the reconstruction and modeling of tracks are also taken into account; these cover track reconstruction efficiency, impact parameter resolution, tracking in dense environments, fake-track rates, and sagitta biases. The impact of the large-$R$-jet $p_T$ resolution uncertainty is evaluated event by event by rerunning the analysis with an additional absolute 2% Gaussian smearing applied to the input jets’ $p_T$ to degrade the nominal resolution. Small-$R$-jet $p_T$ scale and resolution uncertainties are similarly estimated by comparing data with simulation after applying in situ corrections\cite{68}.

Several sources of theoretical uncertainty affecting the signal models were considered. Uncertainties in the matrix element calculation are evaluated by varying the strong coupling constant ($\alpha_s$), the renormalization scale ($\mu_r$), and the factorization scale ($\mu_f$). The effects of uncertainties in the parton distribution functions are evaluated by comparing the $m_{jj}$ distributions obtained when using various alternative PDF sets and taking an envelope of these distributions, as prescribed by the PDF4LHC group\cite{69}. Generator-level variations of the A14 tune’s parameter values are used to cover the initial- and final-state radiation (ISR and FSR, respectively) and multiple parton interaction (MPI) uncertainties. An overall conservative 3% normalization uncertainty is applied to all signals as a result of ISR, FSR, and MPI modeling effects.

**IX. STATISTICAL ANALYSIS**

The statistical framework in the analysis is used to perform hypothesis tests in the SRs, testing the compatibility of the data with both the background-only and signal-plus-background hypotheses. The observable that is fitted is the $m_Y$ distribution of the data in the SR. This fit is repeated several times in overlapping bins of the $X$ candidate mass.

The parameter of interest in the statistical analysis is the signal strength $\mu$, defined as a scale factor multiplying the nominal yield predicted by a 1 pb signal cross section so as to match the observed number of signal events. The background-only hypothesis corresponds to $\mu = 0$. The normalization of the data-driven background estimate is allowed to float, with each normalization factor being fitted independently because the $m_X$ categories overlap. A test statistic based on the profile likelihood ratio calculated in the lowest-order asymptotic approximation is used to test the models corresponding to the signal grid. Systematic uncertainties are included in the fit as nuisance parameters (NPs) with Gaussian constraints. Both the signal strength and all signal systematic uncertainty NPs are correlated across the merged and resolved regions. The significance of an observed excess of data over the background prediction is quantified by the local $p$ value, which is the probability that the background-only model produces an excess at least as large as the one observed.

BumpHunter\cite{70} is used to find excesses in both the anomaly and two-prong SRs, with an algorithm that incorporates only the statistical uncertainty of the data and does not depend on a specific signal shape. It outputs a
\( p \) value that provides a goodness-of-fit metric, along with an interval of the invariant mass corresponding to the largest deviation of the data from the background prediction. It is thus used in all signal regions to search for significant deviations of the data from the background \( m_Y \) distribution. Dedicated studies indicate the fits are dominated by statistical uncertainties, further motivating the discounting of systematic uncertainties in the BumpHunter fits. In the anomaly SR, no subsequent fits are performed using the \( Y \rightarrow XH \) signal model, as this analysis is designed to keep signal-model dependence minimal and the two-prong regions are expected to have higher signal sensitivity across the \( Y \rightarrow XH \) grid. However, given the absence of a significant excess in the two-prong regions, signal-plus-background fits are performed for each \( Y \rightarrow XH \) signal model, and 95\% confidence level (CL) upper limits on the signal cross section are set using a modified frequentist method (CLs) [71].

The background estimation and statistical treatment are validated by comparison with data in the signal-depleted LSB VR for each of the three analysis categories. BumpHunter \( p \) values from all \( m_X \)-bin fits in the VR of the anomaly region approximate a flat distribution between 0 and 1, indicating good background modeling with no systematic biases across the phase space. Figure 7 shows the postfit background prediction compared with the data.
passing the anomaly, merged, and resolved selections, in an example $m_X$ window between 284.5 and 322.5 GeV where the whole kinematic phase space is well populated. Good modeling is observed, and no significant pulls or constraints are observed in the background NPs. In the two-prong regions, checks are made for spurious signals in the LSB through a signal-plus-background fit using each generated signal model. These checks indicate that no significant spurious signal is present beyond those likely to be produced by statistical fluctuations, with no systematic trend across the expected phase space.

**X. RESULTS**

Results of background-only fits of the $m_Y$ distribution across all $m_X$ categories in the anomaly SR show the data to be compatible with the expected backgrounds given their uncertainties. Figure 8 shows the distribution of $p$ values across $m_Y$ and $m_X$ bins, where the background is the result of a background-only fit performed with all statistical and systematic uncertainties, and all uncertainties are included in the $p$-value calculation. The distribution of observed $p$ values is statistically compatible with the expectation, with a $\chi^2$ value per degree of freedom of 1.32 between the distributions.

The largest observed excess in the anomaly SR is in the $m_X$ window [75.5, 95.5] GeV. In this window, the greatest single-bin excess is found for $m_Y$ between 3608 and 3805 GeV. The BumpHunter interval with the greatest significance covers the $m_Y$ range between bin edges of 3424 and 3805 GeV. The $m_Y$ distribution corresponding to this $m_X$ window is shown in Fig. 9, along with the postfit expected background. Studies of the individual jet mass distributions in this $m_X$ category do not reveal any excesses in data that are consistent with a resonant particle decay. Given the number of individual search regions in this analysis, the impact of the trials factor is significant. A calculation is made to determine the global significance of the greatest single-bin deviation, accounting for all the

![Figure 8](image1.png)

**FIG. 8.** The distribution of observed $p$ values across all $m_Y$ and $m_X$ bins in the anomaly signal region, comparing data with the background estimates generated by a background-only fit, displayed in the two-dimensional ($m_X$, $m_Y$) grid. The $p$-value calculation is performed at the center of each $m_X$ bin, and all statistical and background systematic uncertainties are considered. The lowest observed $p$ value corresponds to the bin with $m_Y$ within [3608, 3805] GeV and $m_X$ within [75.5, 95.5] GeV.

![Figure 9](image2.png)

**FIG. 9.** The $m_Y$ distribution associated with the $m_X$ bin [75.5, 95.5] GeV in the anomaly SR, where the background is determined by a background-only fit to data with both the statistical and systematic uncertainties included. The $p$ value and the lower panel of per-bin significances are computed with statistical uncertainties only. The single $m_Y$ bin with the most significant excess is [3608, 3805] GeV. Incorporating the trials factor for all $m_Y$ and overlapping $m_X$ bins in the search, this excess has a global significance of 1.43$\sigma$. 

052009-13
overlapping $m_X$ bins. First, the overlapping bin edges are used to define exclusive, nonoverlapping bins in $m_X$, and an integer is drawn from a Poisson distribution with a mean equal to the background expectation in each exclusive ($m_Y$, $m_X$) bin. Next, this yield is summed across exclusive bins to create a toy estimate for each overlapping bin in which the p value is computed. Finally, this procedure is repeated $N$ times, where $N$ is the number of events, inclusively across all exclusive, nonoverlapping bins. This calculation yields a global significance of $1.43\sigma$ for the excess in the $[3608, 3805]$ GeV bin of $m_Y$.

Results for the two-prong SRs are similarly derived by performing background-only fits and scanning with BumpHunter for incompatibility with data. No significant deviations of data from the predicted background are observed beyond expected statistical fluctuations, in either the merged or resolved SR. An example $m_Y$ distribution in both the merged and resolved SRs for the $m_X$ bin $[284.5, 322.5]$ GeV is shown in Fig. 10, along with the background estimate that is determined from a background-only fit to the data with all statistical and systematic uncertainties included. The ratio of the observed data to the background is shown in the lower panel. The uncertainty band includes both the statistical and systematic effects.

A bilinear interpolation procedure is applied to provide results between fully simulated signal points. The analysis is most sensitive in the very boosted regime, where the $Y$ mass is approximately an order of magnitude larger than the $X$ mass. Sensitivity is lowest in the highly resolved regime, where the required large-$R$ jet reconstruction of the Higgs boson sculpts the signal efficiency to favor high-momentum $X$ particles. The observed limits range from cross sections of $0.341$ fb for the signal point ($m_Y = 5000$ GeV, $m_X = 600$ GeV), to $1.22$ pb for the signal point ($m_Y = 2500$ GeV, $m_X = 2000$ GeV).

The data in the anomaly and two-prong SRs can be used to provide a benchmark comparison of sensitivity across the set of large-$R$ jet topologies considered for the $X$ decay, thereby providing a metric for assessing the level of signal-model dependence in both regions. The 95% CL upper limits on the production cross section of several benchmark signals are calculated for all three signal region selections, by injecting signal into the data until the BumpHunter p value exceeds a significance of $2\sigma$. Seven signal points are considered in this study, including three highly boosted $Y \to XH$ points, one resolved $Y \to XH$ point, and the three alternative jet topologies. Because the systematic uncertainties in the signal efficiency of the anomaly score are not assessed, this comparison is performed using only statistical uncertainties and a postfit background estimate in the limit calculation.

Since the merged region uses $D^2_{trk}$ and thus explicitly tags on the two-prong substructure of the $X$ candidate’s large-$R$ jet in the generated $Y \to XH$ grid, it is possible that this region will outperform the fully unsupervised approach.
FIG. 11. The $p$ value per $m_Y$ bin for both two-prong SRs, calculated using all systematic and statistical uncertainties of the background estimate. Two $m_X$ bins are shown, $[75.5, 95.5]$ and $[113.0, 137.0]$ GeV, which corresponds to a window containing the $W=Z$ and Higgs boson mass, respectively. Events thus fall in (a) a merged $W=Z$ window, (b) a merged Higgs window, (c) a resolved $W=Z$ window, and (d) a resolved Higgs window. The background is determined by a background-only fit to the data with all statistical and systematic uncertainties included. In both $m_X$ windows, the $p$ value has a high approximately constant value for the high $Y$-mass region of the resolved SR (bottom), as this phase-space region is far more likely to produce a highly boosted $J_X$ that falls in the merged SR selection.

FIG. 12. The expected (left) and observed (right) 95% CL limits on the cross section $\sigma(pp \rightarrow Y \rightarrow XH \rightarrow q\bar{q}b\bar{b})$ in picobarns in the two-dimensional space of $m_Y$ versus $m_X$, obtained from a simultaneous fit of both the merged and resolved two-prong signal regions with all statistical and systematic uncertainties. A bilinear interpolation procedure is applied to provide results between fully simulated signal points. The observed limits range from 0.34 fb for the signal point ($m_Y = 5000$ GeV, $m_X = 600$ GeV) to 1.22 pb for the signal point ($m_Y = 2500$ GeV, $m_X = 2000$ GeV). The boundaries of the limit are defined by the simulated signal grid.
While these results cover regions of phase space that have not been studied directly by other searches, some analysis selections overlap with those of other recent ATLAS dijet resonance searches. The $m_X$ bin of [75.5, 95.5] GeV would be sensitive to a $VH$ resonance’s hadronic final state, which is covered by a dedicated analysis using the same dataset [72]. The present approach differs in the tagging techniques used for both the vector boson and Higgs boson but provides a similar 95% CL upper limit on the production cross section of a 3 TeV resonance. Due to its generality, the anomaly SR is expected to be sensitive to the same signatures as the weakly supervised dijet resonance search [13], but a direct comparison is not provided due to the assumptions made here about the mass and decay of the Higgs boson candidates.

XI. CONCLUSION

A search is performed for a heavy new boson $Y$ decaying into a new particle $X$ and a Standard Model Higgs boson in 139 fb$^{-1}$ of 13 TeV $pp$ collision data collected by the ATLAS detector during LHC run 2. The analysis focuses on a fully hadronic final state, where the $X$ particle and $H$ boson are boosted such that their daughter particles are collimated. In the first application of fully unsupervised machine learning to an ATLAS search, a VRNN is trained on jets in data to define an anomaly detection SR, which selects the $X$ particle based solely on its substructural incompatibility with background jets. Two supplementary SRs are designed to separately reconstruct merged and resolved decays of a nominal two-prong $X$ benchmark signal. Sensitivity over the dominant multijet background is enhanced by additional machine-learning applications, namely a NN-based $H \rightarrow b\bar{b}$ tagger and a DNN-based reweighting to ensure good modeling. No significant deviations of the data from the predicted background are observed. The largest excess is found in the anomaly SR with a global significance of 1.43σ when considering all $m_X$ and $m_Y$ bins and is not found to be compatible with the expected signal shape. Results are interpreted as 95% confidence-level upper limits on the cross section $\sigma(pp \rightarrow Y \rightarrow XH \rightarrow q\bar{q}b\bar{b})$, across the two-dimensional space where $m_Y$ is between 1.5 and 6 TeV and $m_X$ is between 65 and 3000 GeV. The most stringent upper limit of 0.341 fb is obtained in the merged SR for the signal point ($m_Y = 5000$ GeV, $m_X = 600$ GeV), while the weakest upper limit of 1.22 pb is obtained for the highly resolved point ($m_Y = 2500$ GeV, $m_X = 2000$ GeV).

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;
Yerevan Physics Institute (YerPhI), Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and Danish Natural Science Council (DNSRC), Denmark; IN2P3-CNRS and CEA-Fundamental Research Division/IRFU = Institute of Research into the Fundamental Laws of the Universe, France; Shota Rustaveli National Science Foundation of Georgia (SRNSFG), Georgia; BMBF, HGF and MPG, Germany; General Secretariat for Research and Innovation (GSRI), Greece; Research Grants Council (RGC) and Hong Kong Special Administrative Region (SAR), China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; Research Council of Norway (RCN), Norway; MEiN, Poland; FCT, Portugal; Ministry of National Education (MNE)/IFA, Romania; Ministry of Education, Science and Technological Development (MESTD), Serbia; Ministry of Education, Science and Sport (MSSR), Slovenia; DSI/INRF, South Africa; MICINN, Spain; Swedish Research Council (SRC) and Wallenberg Foundation, Sweden; State Secretariat for Education, Research and Innovation (SERI), SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; European Cooperation in Science and Technology (COST), ERC, ERDF, Horizon 2020 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 programs cofinanced by EU-ESF and the Greek National Strategic Reference Framework (NSRF), Greece; Binational Science Foundation - National Science Foundation (BSF-NSF) and MINerva, Israel; Norwegian Financial Mechanism 2014-2021, Norway; N CN 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 (Canada), Nordic Grid Facility (NGDF) (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/ GridKA (Germany), INFN-National Center for Frame Analysis (CNAF) (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), Rutherford Appleton Laboratory (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. [73].


[61] F. Chollet et al., Keras (2015), https://keras.io.


ANOMALY DETECTION SEARCH FOR NEW RESONANCES …

PHYS. REV. D 108, 052009 (2023)
<table>
<thead>
<tr>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>İstinye University, Sariyer, Istanbul, Türkiye</td>
<td></td>
</tr>
<tr>
<td>Facultad de Ciencias y Centro de Investigaciones, Universidad Antonio Nariño, Bogotá, Colombia</td>
<td></td>
</tr>
<tr>
<td>Departamento de Física, Universidad Nacional de Colombia, Bogotá, Colombia</td>
<td></td>
</tr>
<tr>
<td>Dipartimento di Fisica e Astronomia A. Righi, Università di Bologna, Bologna, Italy</td>
<td></td>
</tr>
<tr>
<td>INFN Sezione di Bologna, Italy</td>
<td></td>
</tr>
<tr>
<td>Physikalisches Institut, Universität Bonn, Bonn, Germany</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Boston University, Boston Massachusetts, USA</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Brandeis University, Waltham Massachusetts, USA</td>
<td></td>
</tr>
<tr>
<td>Transilvania University of Brasov, Brasov, Romania</td>
<td></td>
</tr>
<tr>
<td>Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Alexandru Ioan Cuza University of Iasi, Iasi, Romania</td>
<td></td>
</tr>
<tr>
<td>National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj-Napoca, Romania</td>
<td></td>
</tr>
<tr>
<td>University Politehnica Bucharest, Bucharest, Romania</td>
<td></td>
</tr>
<tr>
<td>West University in Timisoara, Timisoara, Romania</td>
<td></td>
</tr>
<tr>
<td>Faculty of Physics, University of Bucharest, Bucharest, Romania</td>
<td></td>
</tr>
<tr>
<td>Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovak Republic</td>
<td></td>
</tr>
<tr>
<td>Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic</td>
<td></td>
</tr>
<tr>
<td>Physics Department, Brookhaven National Laboratory, Upton New York, USA</td>
<td></td>
</tr>
<tr>
<td>Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, y CONICET, Instituto de Física de Buenos Aires (IFIBA), Buenos Aires, Argentina</td>
<td></td>
</tr>
<tr>
<td>California State University, California, USA</td>
<td></td>
</tr>
<tr>
<td>Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, University of Cape Town, Cape Town, South Africa</td>
<td></td>
</tr>
<tr>
<td>iThemba Labs, Western Cape, South Africa</td>
<td></td>
</tr>
<tr>
<td>Department of Mechanical Engineering Science, University of Johannesburg, Johannesburg, South Africa</td>
<td></td>
</tr>
<tr>
<td>National Institute of Physics, University of the Philippines Diliman (Philippines), Philippines</td>
<td></td>
</tr>
<tr>
<td>University of South Africa, Department of Physics, Pretoria, South Africa</td>
<td></td>
</tr>
<tr>
<td>University of Zululand, KwaDlangezwa, South Africa</td>
<td></td>
</tr>
<tr>
<td>School of Physics, University of the Witwatersrand, Johannesburg, South Africa</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Carleton University, Ottawa ON, Canada</td>
<td></td>
</tr>
<tr>
<td>Faculté des Sciences Ain Chock, Réseau Universitaire de Physique des Hautes Energies—Université Hassan II, Casablanca, Morocco</td>
<td></td>
</tr>
<tr>
<td>Faculté des Sciences, Université Ibn-Tofail, Kénitra, Morocco</td>
<td></td>
</tr>
<tr>
<td>Faculté des Sciences Semlalia, Université Cadi Ayyad, LPHEA-Marrakech, Morocco</td>
<td></td>
</tr>
<tr>
<td>LPMR, Faculté des Sciences, Université Mohamed Premier, Oujda, Morocco</td>
<td></td>
</tr>
<tr>
<td>Faculté des sciences, Université Mohammed V, Rabat, Morocco</td>
<td></td>
</tr>
<tr>
<td>Institute of Applied Physics, Mohammed VI Polytechnic University, Ben Guerir, Morocco</td>
<td></td>
</tr>
<tr>
<td>CERN, Geneva, Switzerland</td>
<td></td>
</tr>
<tr>
<td>Affiliated with an institute covered by a cooperation agreement with CERN</td>
<td></td>
</tr>
<tr>
<td>Affiliated with an international laboratory covered by a cooperation agreement with CERN</td>
<td></td>
</tr>
<tr>
<td>Enrico Fermi Institute, University of Chicago, Chicago Illinois, USA</td>
<td></td>
</tr>
<tr>
<td>LPC, Université Clermont Auvergne, CNRS/IN2P3, Clermont-Ferrand, France</td>
<td></td>
</tr>
<tr>
<td>Nevis Laboratory, Columbia University, Irvington New York, USA</td>
<td></td>
</tr>
<tr>
<td>Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark</td>
<td></td>
</tr>
<tr>
<td>Dipartimento di Fisica, Università della Calabria, Rende, Italy</td>
<td></td>
</tr>
<tr>
<td>INFN Gruppo Collegato di Cosenza, Laboratori Nazionali di Frascati, Italy</td>
<td></td>
</tr>
<tr>
<td>Physics Department, Southern Methodist University, Dallas Texas, USA</td>
<td></td>
</tr>
<tr>
<td>Physics Department, University of Texas at Dallas, Richardson Texas, USA</td>
<td></td>
</tr>
<tr>
<td>National Centre for Scientific Research “Demokritos”, Agia Paraskevi, Greece</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Stockholm University, Sweden</td>
<td></td>
</tr>
<tr>
<td>Oskar Klein Centre, Stockholm, Sweden</td>
<td></td>
</tr>
<tr>
<td>Deutsches Elektronen-Synchrotron DESY, Hamburg and Zeuthen, Germany</td>
<td></td>
</tr>
<tr>
<td>Fakultät Physik, Technische Universität Dortmund, Dortmund, Germany</td>
<td></td>
</tr>
<tr>
<td>Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, Duke University, Durham North Carolina, USA</td>
<td></td>
</tr>
<tr>
<td>SUPA—School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom</td>
<td></td>
</tr>
<tr>
<td>INFN e Laboratori Nazionali di Frascati, Frascati, Italy</td>
<td></td>
</tr>
</tbody>
</table>
Kyoto University of Education, Kyoto, Japan

Research Center for Advanced Particle Physics and Department of Physics, Kyushu University, Fukuoka, Japan

Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata, Argentina

Physics Department, Lancaster University, Lancaster, United Kingdom

Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom

Department of Experimental Particle Physics, Jožef Stefan Institute and Department of Physics, University of Ljubljana, Ljubljana, Slovenia

School of Physics and Astronomy, Queen Mary University of London, London, United Kingdom

Department of Physics, Royal Holloway University of London, Egham, United Kingdom

Department of Physics and Astronomy, University College London, London, United Kingdom

Louisiana Tech University, Ruston Los Angeles, USA

Fysiska institutionen, Lunds universitet, Lund, Sweden

Departamento de Física Teorica C-15 and CIAFF, Universidad Autónoma de Madrid, Madrid, Spain

Institut für Physik, Universität Mainz, Mainz, Germany

School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom

CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France

Department of Physics, University of Massachusetts, Amherst Massachusetts, USA

Department of Physics and Astronomy, University of New Mexico, Albuquerque New Mexico, USA

Institute for Mathematics, Astrophysics and Particle Physics, Radboud University/Nikhef, Nijmegen, Netherlands

Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amsterdam, Netherlands

Department of Physics, Northern Illinois University, DeKalb Illinois, USA

New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

University of Sharjah, Sharjah, United Arab Emirates

Department of Physics, New York University, New York New York, USA

Ochanomizu University, Otsuka, Bunkyo-ku, Tokyo, Japan

Ohio State University, Columbus Ohio, USA

Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman Oklahoma, USA

Department of Physics, Oklahoma State University, Stillwater Oklahoma, USA

Palacký University, Joint Laboratory of Optics, Olomouc, Czech Republic

Institute for Fundamental Science, University of Oregon, Eugene, Oregon, USA

Graduate School of Science, Osaka University, Osaka, Japan

Department of Physics, University of Oslo, Oslo, Norway

Department of Physics, Oxford University, Oxford, United Kingdom

LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France

Department of Physics, University of Pennsylvania, Philadelphia Pennsylvania, USA

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh Pennsylvania, USA

Laboratório de Instrumentação e Física Experimental de Partículas—LIP, Lisboa, Portugal

Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

Departamento de Física, Universidade de Coimbra, Coimbra, Portugal

Centro de Física Nuclear da Universidade de Lisboa, Lisboa, Portugal

Departamento de Física, Universidade do Minho, Braga, Portugal

Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada (Spain), Spain

Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic

Czech Technical University in Prague, Prague, Czech Republic

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic

Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz California, USA

Departamento de Física, Pontificia Universidad Católica de Chile, Santiago, Chile

Millennium Institute for Subatomic physics at high energy frontier (SAPHIR), Santiago, Chile

Instituto de Investigación Multidisciplinario en Ciencia y Tecnología, y Departamento de Física, Universidad de La Serena, Chile

Universidad Andres Bello, Department of Physics, Santiago, Chile

Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile

Department of Physics, University of Washington, Seattle Washington, USA

Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom

Department of Physics, Shinshu University, Nagano, Japan

Department Physik, Universität Siegen, Siegen, Germany

Department of Physics, Simon Fraser University, Burnaby British Columbia, Canada

Department of Physics, Royal Institute of Technology, Stockholm, Sweden

Departments of Physics and Astronomy, Stony Brook University, Stony Brook New York, USA

Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom

School of Physics, University of Sydney, Sydney, New South Wales, Australia

Institute of Physics, Academia Sinica, Taipei, Taiwan

E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi, Georgia

High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia

Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel

Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece

International Center for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo, Japan

Department of Physics, Tokyo Institute of Technology, Tokyo, Japan

Department of Physics, University of Toronto, Toronto Ontario, Canada

TRIUMF, Vancouver BC, Canada

Department of Physics and Astronomy, York University, Toronto Ontario, Canada

Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Japan

Department of Physics and Astronomy, Tufts University, Medford Massachusetts, USA

United Arab Emirates University, Al Ain, United Arab Emirates

Department of Physics and Astronomy, University of California Irvine, Irvine California, USA

Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden

Department of Physics, University of Illinois, Urbana Illinois, USA

Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia—CSIC, Valencia, Spain

Department of Physics, University of British Columbia, Vancouver British Columbia Canadian Province, Canada

Department of Physics and Astronomy, University of Victoria, Victoria British Columbia Canadian Province, Canada

Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany

Department of Physics, University of Warwick, Coventry, United Kingdom

Waseda University, Tokyo, Japan

Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel

Department of Physics, University of Wisconsin, Madison Wisconsin, USA

Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany

Department of Physics, Yale University, New Haven Connecticut, USA

Deceased.

Also at Department of Physics, King’s College London, London, United Kingdom.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan.

Also at Lawrence Livermore National Laboratory, Livermore California, USA.

Also at TRIUMF, Vancouver British Columbia, Canada.

Also at Department of Physics, University of Thessaly, Greece.

Also at An-Najah National University, Nablus, Palestine.
ANOMALY DETECTION SEARCH FOR NEW RESONANCES …