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

The ATLAS Collaboration

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 $\sqrt{s}$ = 13 TeV pp dataset is used, collected by the ATLAS detector from 2015 to 2018 and corresponding to an integrated luminosity of 139 fb^{-1}. The search targets the high Y mass regime, where the H and X have a significant Lorentz boost. A novel anomaly detection signal region is implemented based on a jet-level score for signal model-independent tagging of the boosted X, 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 to two quarks, which cover topologies where the X is reconstructed as either a single large-radius jet or two small-radius jets. The Higgs is assumed to decay to $b\bar{b}$ and a dedicated neural net-based tagger is used for sensitivity to the boosted heavy flavor topology. No significant excess of data is observed over the expected background, and the results are interpreted in upper limits at 95% confidence level on the production cross section $\sigma(pp \rightarrow Y \rightarrow XH \rightarrow q\bar{q}b\bar{b})$ for signals with m_Y between 1.5 and 6 TeV and m_X between 65 and 3000 GeV.

1 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 2012 discovery of the Higgs boson [1, 2] completed the sequence of particles predicted by the SM. However, many mysteries remain, such as the nature of dark matter and the origin of matter-antimatter asymmetry, which confirm the need for new physics beyond the Standard Model (BSM).

The sensitivity of the Higgs boson mass to radiative corrections implies either extreme fine-tuning in the model or the existence of new physics 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 $O(\text{TeV})$ masses. Because the Higgs boson couples to mass, 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 boson Y, which decays to a Standard Model Higgs H and a new particle X with a mass on the weak scale. A fully hadronic final state is assumed for both particles. Tagging of the boosted Higgs to two b-quarks ($H \to b\bar{b}$ tagging) is applied to enhance the signal using the highest branching fraction decay of the Higgs boson. A novel jet-level implementation of *anomaly detection* based on an unsupervised machine learning architecture is used to select X particles based solely on their incompatibility with the expected SM background. This selection of the X is not strongly dependent on its mass, thus the analysis is sensitive to X masses spanning several orders of magnitude, from $O(10)$ GeV to $O(1)$ TeV. A Feynman diagram for this process can be found in Figure 1, where the X can have a variety of hadronic decays. A nominal benchmark decay of $X \to q\bar{q}$ is generated to provide an interpretation framework for the results. The masses of the parent and daughter particles yield a kinematic scenario where the final state particles are highly Lorentz-boosted, motivating a reconstruction using large-radius (large-R) jets and the use of jet substructure to distinguish the boson decay products. An orthogonal resolved reconstruction is used to recover sensitivity to topologies where the X is less boosted and reconstructed as small-radius (small-R) jets, significantly extending the region of sensitive phase space. As the signal is resonant, it can be detected via a "bump hunt" on the invariant mass of the reconstructed daughter particles.
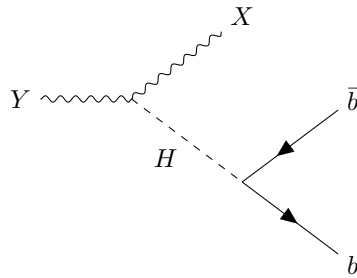

Figure 1: Feynman diagram of the target signal process, where the Y is produced in the initial pp collision and decays to a fully hadronic final state via a SM Higgs boson $H \to b\bar{b}$ and a new particle X. The only assumption applied to the X decay is that it yields a hadronic final state.

Though model-independent in nature, this search is motivated by several key extensions to the Standard Model which predict heavy diboson resonances. Examples of such theories include extended gauge symmetry models [3], warped extra dimensions [4–6], or two Higgs doublet models [7]. Signals are generated with a simplified model based on spin-1 Heavy Vector Triplets (HVT) [8] which reproduces a

large class of explicit BSM models. The HVT phenomenology provides a Lagrangian that fulfills SM 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 performed by ATLAS using 36.1 fb^{-1} under the assumption of $X \rightarrow q\bar{q}$, with no significant excess found covering Y masses from 1 to 4 TeV and X masses from 50 to 1000 GeV [9]. In addition to the increased luminosity of the dataset, this result includes several key improvements with respect to this last iteration, such as a neural net-based tagger optimized for the boosted $H \rightarrow b\bar{b}$ topology, anomaly detection for enhanced signal model independence, and the usage of two orthogonal regions to capture both boosted and resolved reconstruction of the nominal X decay to two quarks. The CMS experiment performed a multidimensional diboson resonance search with the full Run 2 dataset of 139 fb^{-1} that includes sensitivity to the signature discussed here [10]. A maximum local (global) significance of 3.6 (2.3) standard deviations was observed for two mild excesses of events, at masses of 2.1 and 2.9 TeV.

The analysis regions are built by selecting two large-R jets with additional criteria to enrich the presence of Higgs and X particles. Three overlapping analysis categories are defined based on 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 that have different large-R jet decay topologies for the X, in addition to the two light quark final state that is used across the generated (m_Y, m_X) signal grid. The background to the signal process is composed primarily of jets from quantum chromodynamics (QCD) processes. It is estimated with a fully data-driven method that incorporates a deep neural net (DNN)-based reweighting to ensure good modeling. The final fit is performed on the reconstructed invariant mass of the Y in overlapping categories of the X candidate mass to further enrich the signal-to-background ratio. Results are presented as limits on the cross section times branching ratio of the generic HVT process.

2 ATLAS detector

The ATLAS detector [11] is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4π coverage in solid angle[1]. 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 [12, 13]. The next layer 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/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap 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

---

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The $x-$axis points from the IP to the centre of the LHC ring, and $y-$axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -ln\, tan(\theta/2)$.

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. An extensive software suite [14] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Data and Monte Carlo simulation

This search is performed with LHC pp collision data collected by the ATLAS detector during all years of Run 2 from 2015 to 2018. After the application of data quality requirements [15] that ensure good working condition of all detector components, the dataset corresponds to an integrated luminosity of 139.0 ± 2.4 fb^{-1} [16, 17]. The recording of data events is triggered by the presence of high-p_T large-R jets, built at trigger-level using calorimeter-cell energy clusters calibrated to the hadronic scale utilising the local cell signal weighting method [18]. The single large-R jet trigger with the lowest threshold is used across data-taking years, corresponding to a requirement on the trigger jet to have $p_T \geq 360$ GeV (2015), $\geq$ 420 GeV (2016), and ≥ 440 GeV (2017 and 2018). Further offline selection criteria have been imposed (see Section 4) which ensure that selected events are all in the kinematic regime where the trigger is fully efficient, avoiding the effect of the trigger turn-on in the analysis.

Monte Carlo (MC) event generators are used to simulate the signal targeted by this search. The generation of the simulated event samples includes the effect of *pile-up*, defined as the mean number of interactions per bunch crossing. NNPDF23LO PDF [19] and the ATLAS A14 [20] tune to underlying-event data were used. The effect is assessed with the inclusion of overlaid minimum-bias events, as well as the effect on the detector response due to interactions from bunch crossings before or after the one containing the hard interaction. Events in the simulation are weighted by data-taking period in order to reproduce the observed pile-up distribution. The detector response is simulated with a GEANT4 [21] based framework [22] and the events were processed with the same reconstruction software as that used for data.

A generic HVT [8, 23] model produced via qq scattering is used as a baseline for the $Y \rightarrow XH$ signal samples. The generation of the $W' \rightarrow WH$ process is modified to replace the W with a new spin-1 boson X with a width of 2 GeV and a 100% branching ratio to $u\bar{d}$. The Higgs boson is generated with a nominal mass of 125.5 GeV and decays only to $b\bar{b}$. Samples with resonance Y masses between 1-6 TeV are generated using MADGRAPH5 [24] 2.7.2 and 2.7.3 interfaced to PYTHIA8.2 [25] for shower and hadronization with NNPDF23LO PDF [19] and the ATLAS A14 [20] tune to underlying-event data. A two-dimensional grid is generated with 195 signal points defined by a Y and X mass value, with Y masses in the range of 1.5 to 6 TeV and X masses in the range of 65 to 3000 GeV.

In addition to the $Y \rightarrow XH$ signal grid, three signals with different jet topologies were used to craft and assess the degree of model independence in the analysis sensitivity. They are all generated with PYTHIA8 using the same tune (A14) and PDFs (NNPDF23LO) as the $Y \rightarrow XH$ signal grid. The HVT $W' \rightarrow WZ$ configuration from Ref. [26] was used to create generic resonance signatures of the form $A \rightarrow BC$, where the daughter particles decay to either three light quarks or two heavy flavor quarks. A benchmark choice of the parent A mass at 3 TeV and daughter masses at 200 and 400 GeV creates the same kinematic scenario as the $Y \rightarrow XH$ phase space of interest, leading to boosted decays that are realized as large-R jets. Depending on the daughter decay to either qqq or $b\bar{b}$, the large-R jet is produced with three-prong substructure or displaced vertices, respectively. Lastly, the PYTHIA8 Hidden Valley Model A configuration is used to

generate a dark jet signal that arises from the decay of a Z' dark sector mediator to two dark quarks, where the Z' has the same benchmark mass of 3 TeV [27]. 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 distinguish using traditional substructure variables that focus on specific signal models, making them an ideal target for anomaly detection based on background-only characterization.

4 Event selection

The experimental signature of the $Y \rightarrow XH$ signal contains at least two jets with high transverse momentum (p_T). 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 based on the selection criteria of the X. Each signal region utilizes five background estimation regions, composed of three control regions (CRs) and two validation regions (VRs), that are defined for each of the three SRs using common selections on the Higgs boson. The definitions of these regions and motivation for the selection criteria are provided in this section.

4.1 Object selection

Jets are built with FastJet [28] to cluster constituent detector signals using the anti-k_t algorithm [29] with two radius parameters: $R = 1.0$ for the large-R jets J, and $R = 0.4$ for the small-R jets j. Large-R jets must satisfy $p_T > 200$ GeV and $|\eta| < 2.0$. They are constructed from a combination of tracks and calibrated calorimeter energy clusters known as Track-CaloCluster (TCC) objects [30]. Such objects take advantage of the excellent energy resolution of the ATLAS calorimeter, while incorporating the angular resolution capability of the tracker at very high energy where the calorimeter is unable to resolve the needed structure within the jet. This is particular advantageous for measuring substructure in highly boosted jets, as is necessary in this search to distinguish $Y \rightarrow XH$ signal from multijet background. To minimize the impact of pile-up, they are trimmed [31] by removing any $R = 0.2$ subjets with less than 5% of the p_T of the associated large-R jet. The jet energy and mass is calibrated with an MC-based method [32].

Small-R jets are built from particle flow constituents [33], which provide improved accuracy on the charged hadron energy measurement through measurements of momenta from the ID. They are required to pass $p_T > 20$ GeV and $|\eta| < 4.5$. An additional event-level veto is applied to reject events with a jet that is likely to come from calorimeter noise, beam-induced background, or cosmic rays [34].

Leptons are not explicitly used in the analysis, but are included in an overlap removal procedure that prohibits the double-counting of overlapping objects in an event. Electrons are required to have $p_T > 7$ GeV and $|\eta| < 2.47$. Muons are required to have $p_T > 7$ GeV and $|\eta| < 2.7$. To ensure that leptons originate from the interaction point, transverse and longitudinal track impact parameter criteria, with respect to the beam line, are imposed, $|d_0/\sigma(d_0)| < 3(5)$ for muons (electrons) and $|z_0 \sin\theta| < 0.5$ mm respectively. 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 nearby electrons and 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 nearby muons

and 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 the large-R jet, the large-R jet is removed if the separation between the electron and the large-R jet satisfies $\Delta R < 1.0$.

4.2 Analysis regions

The analysis relies on the reconstruction of charged particles with $p_T > 500$ MeV in the inner detector to reconstruct pp collision vertices, where the primary vertex is chosen to be the vertex with the largest $\Sigma p_T{}^2$ for the tracks associated with the vertex. The recording of data events is triggered by the presence of a single high-p_T J. Selections on the invariant mass of the two large-R jets $m_{JJ} > 1.3$ TeV and the transverse momentum of the leading large-R jet $p_T(J_1) > 500$ GeV are imposed to ensure the trigger is fully efficient. Only the two highest p_T large-R jets in an event are considered, and at least one of their masses is required to be > 50 GeV.

The primary SR based on anomaly detection is referred to as the *anomaly SR*. The remaining two SRs target the benchmark $X \rightarrow q\bar{q}$ decay and are thus called the *two-prong SRs*, differing based on reconstruction of the X 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 are non-overlapping with one another. The analysis CRs include the high sideband (HSB) of the Higgs candidate mass, 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 candidate mass, which validate the reweighting derived from the HSBs. The full analysis flow is visualized in Figure 2, with selection details provided below.

As the SRs are constructed by selecting on the different properties of the X and H jets, an ambiguity resolution is required to determine which of the two J in the event is more likely to be the Higgs boson and thus subject to the Higgs boson tagging selection criteria. This is done using a neural net-based classifier which separates bosons decaying to $b\bar{b}$ from top quark and QCD jets [35]. The tagger trains over the large-R jet p_T and η, along with the subjet flavor tagging score DL1r [36] for up to three track subjets constructed with a p_T-dependent variable radius [37]. The tagger version used here includes a reweighting of all training inputs to have the same p_T and η distributions, to minimize bias of the tagger to high-p_T or central jets respectively. The outputs of the NN are three classification scores corresponding to the likelihood of the jet originating from a Higgs boson (p_{Higgs}), top quark (p_{top}), or multijet process (p_{multijet}), which are subsequently combined into the jet-level discriminant $D_{H_{bb}}$ as shown in Equation 1. Here f_{top} determines the weight of the top background shape in the final discriminant, and a value of 0.25 is chosen based on signal-to-noise optimization studies.

$$D_{H_{bb}} = \ln \frac{p_{\text{Higgs}}}{f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{\text{top}}) \cdot p_{\text{multijet}}} \tag{1}$$

To perform the large-R jet ambiguity resolution, $D_{H_{bb}}$ is computed for both J in the event. The jet with the larger value of $D_{H_{bb}}$ is labeled as the Higgs candidate (J_H), and the other J is by default the X candidate (J_X), therefore determining which jet is subject to further H and X tagging. This procedure is $> 90\%$ accurate across the highly boosted regime of the signal grid ($m_X/m_Y < 0.5$), dropping for less boosted points where the boosted $H \rightarrow b\bar{b}$ tagging is less efficient. The resulting distribution of $D_{H_{bb}}$ for the J chosen as the Higgs candidate, for both data and representative $Y \rightarrow XH$ signals at preselection, is shown
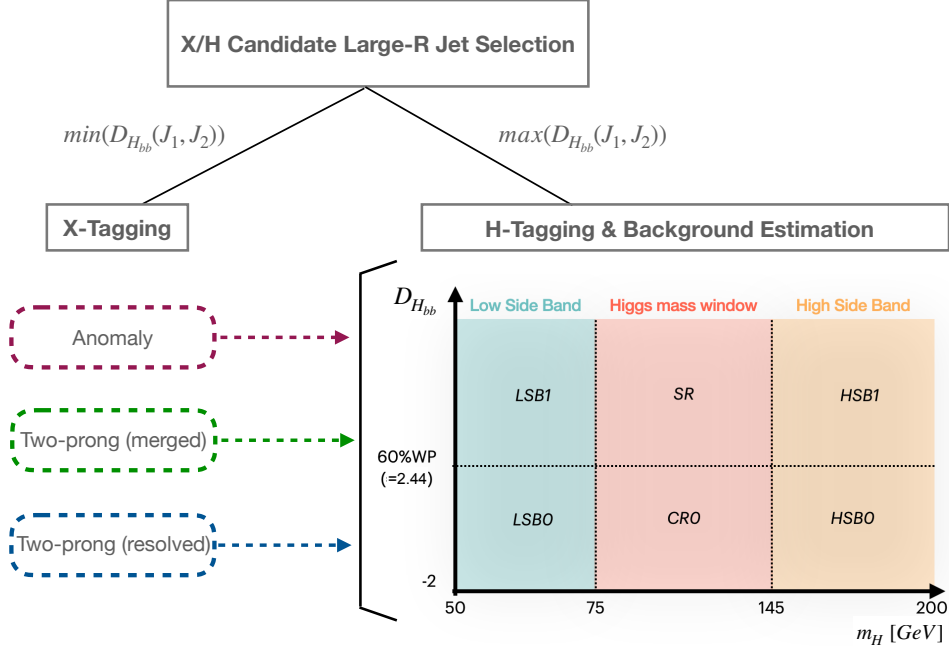
Figure 2: Illustration of the selection flow after preselection and the analysis regions of the $Y \to XH$ search. The first step of resolving the ambiguity between the two large-R jets into X and H candidates dictates which jet is used to make the first selection on either D_2^{trk} or anomaly score of the X candidate. This sorts the preselection events into three separate categories, namely anomaly, two-prong merged, and two-prong resolved. Each region has its own background estimation and validation regions. The CRs (CR0, HSB0, HSB1), VRs (LSB0, LSB1), and SR are defined in the same way for all three analyses using the H candidate mass and $D_{H_{bb}}$ score. $D_{H_{bb}}$ is bounded at the lower edge by the skimming selection requiring J_H to have $D_{H_{bb}} > -2$.

in Figure 3. Additionally, a preselection of $D_{H_{bb}} > -2$ is applied to remove events that are determined to be not $H \to b\bar{b}$ -like, thus ensuring the data-driven background estimation focuses on an area of phase space that is close to that of the signal.

The anomaly detection region is defined through a selection on the jet-level anomaly score (AS) of the X candidate [38]. The AS is defined using a variational recurrent neural network (VRNN) [39] trained over large-R TCC jets in the ATLAS Run 2 dataset satisfying the trigger plateau criteria, the jet requirements described in Section 4.1, and p_T (J) >1.2 TeV. The p_T selection is designed to restrict input jets to highly boosted topologies, which are both well-described by the k_t-sorted sequence modeling and the most difficult to distinguish with regular substructure methods. 90% of the data is used for training and 10% for validation. As 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 according to a signal or background categorization.

The input jets are modeled as a sequence of up to twenty constituent four-vectors per jet, ordered in k_t splitting starting from the highest p_T constituent. The training is also conditioned over four high-level variables, namely the energy correlator substructure variable D_2 [40] for two-prong sensitivity, the N-subjettiness ratio τ_{32} [41] for three-prong sensitivity, and the two k_t-splitting scale ratios d_{12} and d_{23} [42]. 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
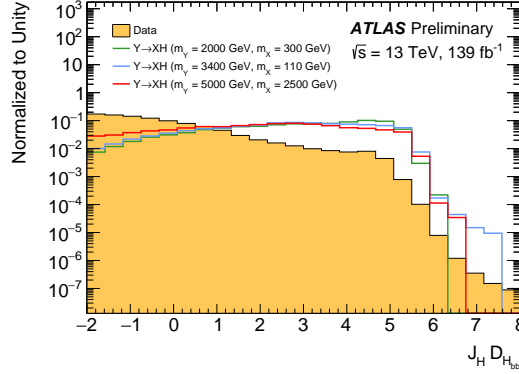
Figure 3: Distributions of the J_H candidate $D_{H_{bb}}$ score in data after preselection requirements are applied. Also shown are three $Y \to XH$ simulated signals, labelled by the masses of the Y and X particles. A selection of $D_{H_{bb}} >$ 2.44 defines a working point that is 60% efficient for the selection of the boosted $H \to b\bar{b}$ topology across the full p_T range. All distributions are normalized to unity.

jets originating from QCD processes. An alignment procedure is applied to each jet that rescales to the same p_T, boosts to the same energy, and rotates to the same orientation in η and ϕ, minimizing the ability of the VRNN to tag only on anomalous kinematic properties without considering internal constituent properties.

The loss function of the VRNN is composed of two terms: a reconstruction error term, to minimize differences of the decoded result with respect to the original input, and a Kullback-Leibler (KL) divergence term [43] from the Gaussian prior distribution to the encoded approximate posterior distribution. The sum of these terms corresponds to the negative of the evidence lower bound. As it is minimized, the KL divergence from the true posterior distribution to the encoded posterior distribution is jointly minimized, ensuring that the architecture accurately describes the underlying distribution of data [44]. The AS is therefore defined as a function of the KL divergence term, such that more anomalous jets populate higher values and background populates lower values.

Figure 4 shows the resulting distribution of the X candidate AS after analysis preselection is applied, comparing data to three example $Y \to XH$ signals as well as the three signals with alternative X decays. The absence of a signal model in the construction of the score means that it is not expected to outperform analytical or supervised methods for any single signal model. However, the approach of using the VRNN to characterize the background distribution provides broad sensitivity to a variety of signals that differ from jets originating due to QCD processes. The most notable difference in AS shape is found for the 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 AS distribution for the highly resolved point with (m_Y= 5000 GeV, m_X= 2500 GeV) populates lower and thus less anomalous values than the data, making this signal indistinguishable by the AS approach. A selection of AS > 0.5 is chosen for the J_X candidate that provides modest but comparable discrimination power over data for the disparate signal jet topologies considered across the full anticipated p_T range of the X candidate jets. The cut enhances the signal-to-background ratio for the two boosted points, (m_Y= 2000 GeV, m_X= 300 GeV) and (m_Y= 3400 GeV, m_X= 110 GeV), by approximately 25% with respect to the inclusive selection.

Regions focusing on the benchmark $X \to q\bar{q}$ decay are used to supplement the anomaly detection analysis and interpret the results. The merged region is defined applying a selection of $D_2^{trk} < 1.2$, where D_2^{trk} is
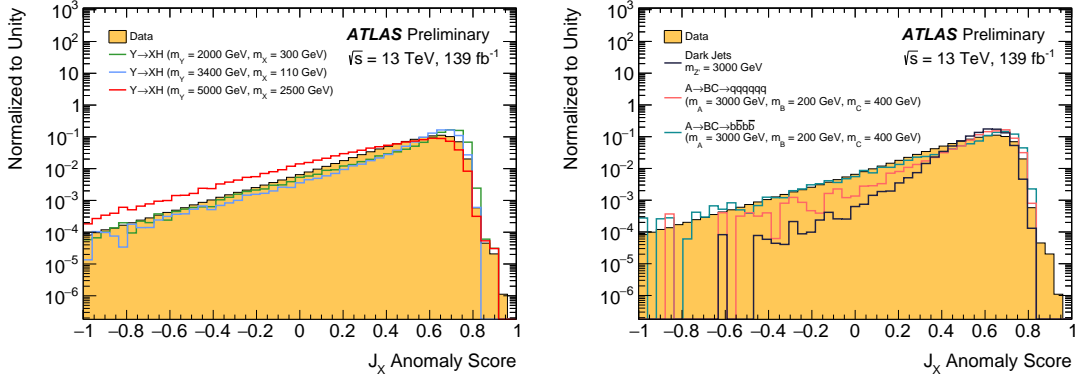
Figure 4: Distributions of the J_X candidate anomaly score (AS) in data after preselection requirements are applied. Also shown are three $Y \to XH$ simulated signals (left), labelled by the masses of the Y and X particles, and the three additional signals with alternative X decay hypotheses, namely heavy flavor, three-prong, and dark jet (right). A selection of AS > 0.5 on the J_X defines the anomaly analysis category. All distributions are normalized to unity.

the energy correlator variable D_2 computed using only tracks associated to 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^{trk} with the standard D_2 variable at analysis preselection have shown good compatibility between D_2^{trk} and D_2, ensuring its sensitivity to two-prong signals with respect to multijet background. As jets with two-prong substructure have lower values of D_2^{trk}, this selection enhances the presence of fully merged decays where the X decay products are well-contained by a single J. The value of 1.2 is chosen to maximize signal-to-background discrimination across the p_T of the X candidate for all mass points in the $Y \to XH$ grid. Distributions of D_2^{trk} for J_X are shown in Figure 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 showed poor reconstruction of the X mass by the large-R jet. To achieve better sensitivity in this regime, a resolved selection is defined requiring D_2^{trk} of J_X to be greater than 1.2, corresponding to a less boosted X 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, and discarding the two with the smallest ΔR to the Higgs boson candidate J_H. m_Y is then computed using the Higgs boson candidate large-R jet and the two small-R jets associated to the X. Additional cuts of $|\Delta y_{j1,j2}| < 2.5$ and $p_T^{bal} < 0.8$ are applied to aid accurate resolved X reconstruction, where y refers to rapidity and p_T^{bal} is defined as the difference of the X small-R jet transverse momenta divided by their sum. The resolved region significantly recovers signal efficiency for the points with higher m_X/m_Y, corresponding to less boost for the X candidate decay.

The Higgs boson tagging is performed after the J_X selection is applied to sort the analysis into the three categories for which a background estimation is derived. For all signal regions, a working point that provides a flat 60% efficiency across jet p_T is applied ($D_{H_{bb}} > 2.44$) to the Higgs boson candidate J_H, along with a mass window requirement of 75 < m_H < 145 GeV. The data-driven background estimation is developed using events from the HSB of the Higgs boson candidate jet (145 GeV < m_H < 200 GeV), which are further classified into HSB0 (HSB1) if the Higgs boson J fails (passes) the $D_{H_{bb}}$ selection. Validation is performed in the LSB, where the reconstructed Higgs boson mass is required to be between 65 and 75 GeV. LSB0 and LSB1 are similarly defined as having a Higgs boson candidate that fails or passes the $D_{H_{bb}}$ tagging criterion, respectively. CR0 is defined as the set of events where the H is in the SR mass window but fails the $D_{H_{bb}}$ tagging, to which the DNN-based reweighting is applied to generate the final
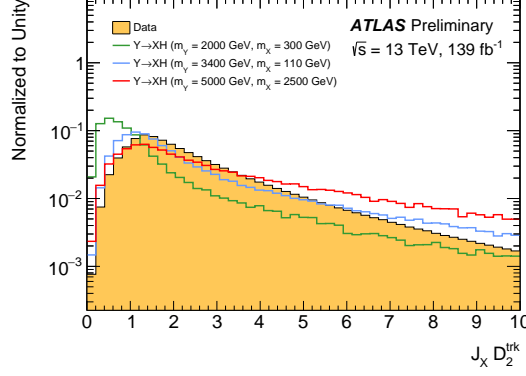
Figure 5: Distributions of J_X candidate D_2^{trk} in data after preselection requirements are applied. Also shown are three $Y \rightarrow XH$ simulated signals, labelled by the masses of the Y and X particles. A value of $D_2^{trk} = 1.2$ therefore separates the analysis events into merged (< 1.2) and resolved (> 1.2) two-prong categories. All distributions are normalized to unity.

background prediction in the SR.

In total, three SRs and fifteen background estimation regions are used in the analysis. A summary of all region definitions can be found in Table 1.

Parameter	Preselection requirements			
m_{JJ} [GeV]	> 1300			
$p_T(J_1)$ [GeV]	> 500			
m_J [GeV]	$m_{J_1} > 50 \parallel m_{J_2} > 50$			
$D_{H_{bb}}$	> -2			
	Signal regions			
	Merged	Resolved	Anomaly	
m_H [GeV]	(75, 145)			
$D_{H_{bb}}$	> 2.44			
D_2^{trk}	< 1.2	> 1.2	-	
$\|\Delta y_{j_1,j_2}\|$	-	< 2.5	-	
p_T^{bal}	-	< 0.8	-	
Anomaly Score	-	-	> 0.5	
	Background estimation regions			
	CR0	HSB0	HSB1	LSB0	LSB1
m_H [GeV]	(75, 145)	(145, 200)		(65, 75)	
$D_{H_{bb}}$	< 2.44	< 2.44	> 2.44	< 2.44	> 2.44

Table 1: Preselection requirements, as well as optimized requirements defining the SRs and background estimation regions.

The signal search in the two-dimensional space of m_Y versus m_X employs sliding windows of the J_X candidate mass spectrum, dividing the data into a series of overlapping m_X ranges for which the m_{JJ} distribution is fit. The choice of binning for m_Y and m_X is chosen based on the expected signal mass resolution, with modifications to account for limited statistics in data. In the m_Y spectrum, bins are widened

at higher values of m_Y to ensure that at least one event is available in each bin using LSB1 data. For the m_X categories, an initial set of mass windows is generated from a linear fit of m_X resolution to m_X value. The first window is chosen to have a bin center of m_X = 65 GeV, the lowest generated signal X mass, and its width chosen to be twice the resolution obtained from the fit. Subsequent windows are generated in the same way, with their bin centers chosen to be higher than the previous window's bin center by a value equal to half of that window's resolution. The edges of high-mass windows are expanded symmetrically based on statistics in LSB1, until at least 10 events are present in each final m_X bin. In order to protect against duplicate windows, all resulting windows which share at least one bin edge are replaced by a single window encompassing their union. Finally, to accommodate the highest generated signal mass point at m_X = 3000 GeV, the final bin's high edge is set to a value of 3200 GeV.

5 Background estimation

The background in the SRs mainly arises from high p_T multijet events. Simulation for such QCD processes includes well-known mismodelings, and is computationally expensive to generate. Therefore in this analysis, the background estimation is fully data-driven and derived from regions that are orthogonal to the SR based on the Higgs boson jet criteria.

The shape of the expected m_{JJ} distribution in the SR is obtained from data in the CR0, and weights are derived that can be applied to HSB0 to reproduce the shape found in HSB1. This procedure is validated by applying the weights to data in LSB0 and comparing the resulting m_{JJ} spectrum to that observed in LSB1 data. The generation of weights is performed inclusively in the X candidate selection, and applied separately to create three background predictions, one for each SR.

The reweighting function is defined as the ratio of the multi-dimensional probability distribution functions (PDFs) of the data in HSB1 to data in HSB0. In this analysis, the statistical procedure of *direct importance estimation* [45] 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 the loss function is minimized to produce weights that can accurately reproduce the observed ratio in data [46].

The DNN is built using a fully connected sequential model from KERAS [47] with 3 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 is randomly truncated ("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 [49] as backend. 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.

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

The DNN outputs event-level weights that are assumed approximately independent of m_H and can be applied to an untagged region to produce the m_{JJ} shape in the corresponding $H \rightarrow b\bar{b}$ -tagged region.

These weights are validated using data from 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 region. Three curves are shown in total, comparing the LSB0 data before and after DNN reweighting is applied to the target data distribution in LSB1. These variables are chosen to focus on kinematic variables over which the background estimation is extrapolated to generate the SR prediction. Good agreement is observed of the reweighted shapes to the true tagged data in all distributions, suggesting a robust background model. As the training is performed inclusively of the X-tagging, the same conclusion holds for the anomaly and two-prong resolved LSB regions. Further, as the reweighting is applied to m_{JJ} distributions after the X tagging selection is applied, no extrapolation across D_2^{trk} or AS is required, and the method shows similar background modeling across all three SRs. Any minor residual differences between predicted background and data are covered in the SR by the non-closure systematic uncertainty, described in Section 6.
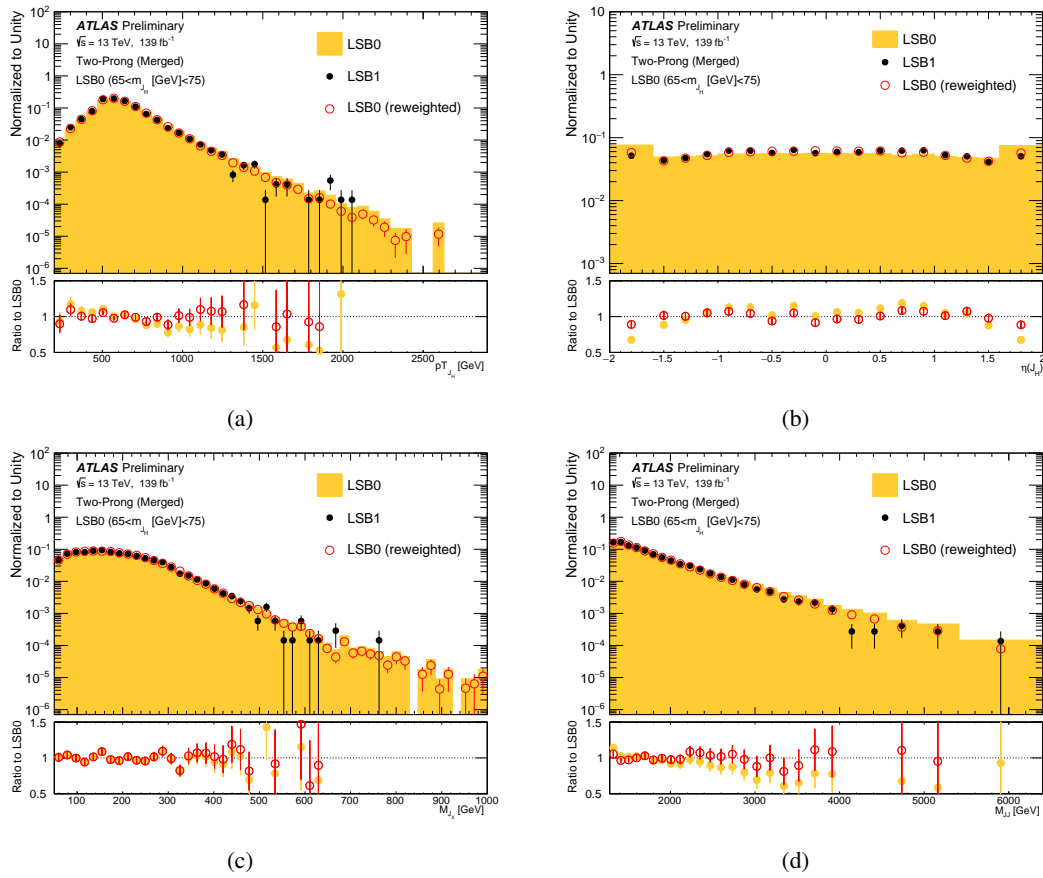

Figure 6: Distribution of the H candidate J p_T (a) and η (b), along with the mass of J_X (c) and m_{JJ} (d) in the merged LSB validation region, overlaying data from LSB1 with the data in LSB0 shown before (orange) and after (red) reweighting. The ratio of the LSB1 data to both the LSB0 data (orange) and the reweighted LSB0 data (red) is shown in the lower panel. Error bars indicate statistical uncertainties.

6 Systematic uncertainties

Systematic uncertainties are applied to both the data-driven background and the simulated signal. All background uncertainties are derived using an m_{JJ} shape that is inclusive in m_X, and applied to each

m_X category fully correlated across m_{JJ} bins. Variations on the background shape are derived in a custom way from three sources.

The first is on training the DNN in an alternate region of $165 < m_H < 200$ GeV, to account for potential variations in obtained weights due to differences in phase space. This region has approximately the same statistics and tagging efficiency as the nominal training region, helping to isolate the effect of the particular DNN model on the obtained reweights. Up and down variations are defined by symmetrizing the shape difference in m_{JJ} between the two different models, creating an effect of $O(1-10)\%$ across the distribution.

Another DNN variation is built to account for the finite statistics of the training sample and the random initialization of the weights. It is estimated with a bootstrap procedure [46] where a set of 100 bootstrap networks are trained, each time varying the training dataset by re-sampling it with replacement. Two additional templates are then defined with the median weight for each event, plus or minus half of the interquartile (IQR) range, defining the upper and lower symmetric error bands. This corresponds to a $O(1)\%$ effect across m_Y.

Lastly, a non-closure uncertainty is included to cover modeling discrepancies that may arise from extrapolating weights derived from the NN training in the HSB 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. In order to not be sensitive to statistical fluctuations in the LSB, a smoothing is applied to the variation where it is rebinned to reduce the relative statistical uncertainty. The non-closure is negligible for low m_{JJ} and rises to $O(10)\%$ in the m_{JJ} tails.

Both normalization and shape uncertainties are applied to the simulated signals. A flat uncertainty of 1.7% is applied for the luminosity as computed by the LUCID-2 detector [50]. The uncertainty from the trigger selection is negligible, as the requirement on $m_{JJ} > 1.3$ TeV ensures that the trigger is fully efficient, and thus not included.

Scale factors (SFs) computed to match the $H \rightarrow b\bar{b}$ tagger efficiency between data and simulation are included that scale the signal template, and their uncertainties are similarly propagated to the signal normalization. These SFs are computed using the methodology from Ref. [51], with the substitution of an updated $D_{H_{bb}}$ that includes the η-reweighting of inputs as described in Section 4.2. They are binned in large-R jet p_T, where the highest p_T bin SF is extrapolated to cover the upper end of the p_T regime probed by the analysis selection. Data and MC agreement for these SFs is verified with $Z \rightarrow b\bar{b}$ events in the context of the calibration analysis. 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 systematics arise from uncertainty on the jet scale and resolution, for both 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 track-to-calorimeter double ratios between the data and the MC, where any observed differences are assigned as baseline systematic uncertainties [52]. Even though the analysis relies on jets built from TCC objects, 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 vs. track-based p_T by cross-calibrating between per-jet TCC and calorimeter p_T and found it to be negligible [26]. The size of the total jet p_T scale uncertainty varies with p_T and η and is typically around $\pm 5\%$. Additional uncertainties due to the reconstruction and modelling of tracks are taken into account as well, which cover track reconstruction efficiency, impact parameter resolution, tracking in dense environments, track fake rate, 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 scale and resolution uncertainties are similarly estimated through data to MC comparisons and in-situ corrections [53].

Several sources of theoretical uncertainty affecting the signal models were considered: Uncertainties in the matrix element calculation are evaluated varying the strong coupling constant (α_S), the renormalization (μ_R) and the factorization scale (μ_F). Uncertainties on the behavior of the parton distribution functions are evaluated by comparing the m_{JJ} distributions for various alternate PDF sets and taking an envelope of the resulting distributions, as prescribed by the PDF4LHC group [54]. Generator-level variations of the A14 tuning parameters are used to cover the initial and final state radiation (ISR, FSR) and multiple parton interaction (MPI) uncertainties. An overall conservative 3% normalization uncertainty is applied to all signals as a result of ISR/FSR/MPI modeling effects.

7 Statistical analysis

The statistical framework in the analysis is used to perform hypothesis testing in the SRs, for the compatibility of the data with both the background-only and the signal-plus-background hypotheses. The observable that is fit is the m_{JJ} 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 μ, defined as a scale factor on the total number of signal events with respect to the nominal yield predicted by an assumption of a 1 pb signal cross section σ. The background only hypothesis corresponds to $\mu = 0$, and the hypothesis of the full signal plus background gives $\mu = 1$. The normalization of the data-driven background estimation is allowed to float, with each normalization factor being fit independently as the m_X categories are overlapping. A test statistic based on the profile likelihood ratio using the lowest order asymptotic approximation is used to test the models proposed by the signal grid. Systematic uncertainties are incorporated into the fit as nuisance parameters (NPs) with Gaussian constraints. Both the signal strength and all signal systematic NPs are correlated across merged and resolved regions. The significance of an excess observed in data over the background prediction is quantified by the local p_0, which is the probability of the background only model to produce an excess at least as large as the one observed.

BumpHunter [55] is used to find excesses in both the anomaly and two-prong SRs, with an algorithm that incorporates only statistical uncertainty on 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 data from background. In the anomaly SR, no fits are performed that use 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 give stronger sensitivity to the $Y \rightarrow XH$ grid. However, given an absence of significant excess in the two-prong regions, signal-plus-background fits are performed for each $Y \rightarrow XH$ signal model, and 95% confidence level upper limits on the signal cross section are set using a modified frequentist method (CL$_s$) [56].

The background estimation and statistical treatment are validated to 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 post-fit background prediction compared to the data for anomaly, merged, and resolved selections, in an example m_X window between 284.5 and

322.5 GeV where all kinematic regimes are well-populated. Good modeling is observed, and no significant pulls or constraints are observed in the background NPs. In the two-prong regions, the spurious signal is checked in the LSB through signal-plus-background fit using each generated signal model. The result of this check indicates no significant spurious signal is present outside of what can be produced by statistical fluctuations, with no systematic trend across the expected phase space.

8 Results

Results of background-only fits of the m_{JJ} distribution across all m_X categories in the anomaly SR show good compatibility of data to expected background, after incorporating all statistical and systematic uncertainties. The largest deviation is in the m_X window [75.5, 95.5] GeV, where the BumpHunter interval covers the m_Y range between bin edges of 3424 and 3805 GeV. The excess corresponds to a p-value of 9.1×10^{-3} considering only statistical uncertainties. The m_{JJ} distribution corresponding to this window is shown in Figure 8, along with the post-fit expected background. Studies of the individual jet mass distributions in this m_X category do not reveal any excesses in data that can be 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 this deviation accounting for all the overlapping m_X bins, where the overlapping bin edges are used to define exclusive, non-overlapping bins in m_X; an integer is drawn from a Poisson distribution with a mean of the background expectation in each exclusive (m_Y, m_X) bin; this yield is summed across exclusive bins to create a toy estimate for each overlapping bin in which the p-value is computed; and this procedure is repeated N times where N is the number of events inclusively across all exclusive, non-overlapping bins. The trials factor is then the fraction of such pseudo-experiments where the largest observed excess is greater than that observed in the single exclusive (m_Y, m_X) bin. This calculation yields a global significance of 1.47σ for this excess.

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 are observed with respect to the predicted background beyond expected statistical fluctuations, in either the merged or resolved SR. An example m_{JJ} distribution in both the merged and resolved SRs for the m_X bin [284.5, 322.5] GeV is shown in Figure 9, along with the background estimation that is determined from a background-only fit accounting for all uncertainties. Figure 10 shows a summary of the per-bin p-values in each m_Y bin for selected m_X bins that are centered on key X candidate mass hypotheses, namely at the W, Z, and H boson masses.

Given the absence of significant excesses in the data, signal-plus-background fits are performed to determine the 95% CL upper limit on the cross section of the $Y \rightarrow XH$ process. A summary of the expected and observed limits in the 2D grid of the $Y \rightarrow XH$ signals is given in Figure 11, combining results from the merged and resolved regions. A bilinear interpolation procedure is applied to provide results in 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 worst in the highly resolved regime, due to the required large-R J reconstruction of the Higgs boson which sculpts signal efficiency to high momentum X particles. The observed limits range from cross sections of 0.342 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 decays considered for the X, thereby providing a metric for assessing the level of signal model-dependence in both regions. As the anomaly SR is mostly sensitive to highly boosted
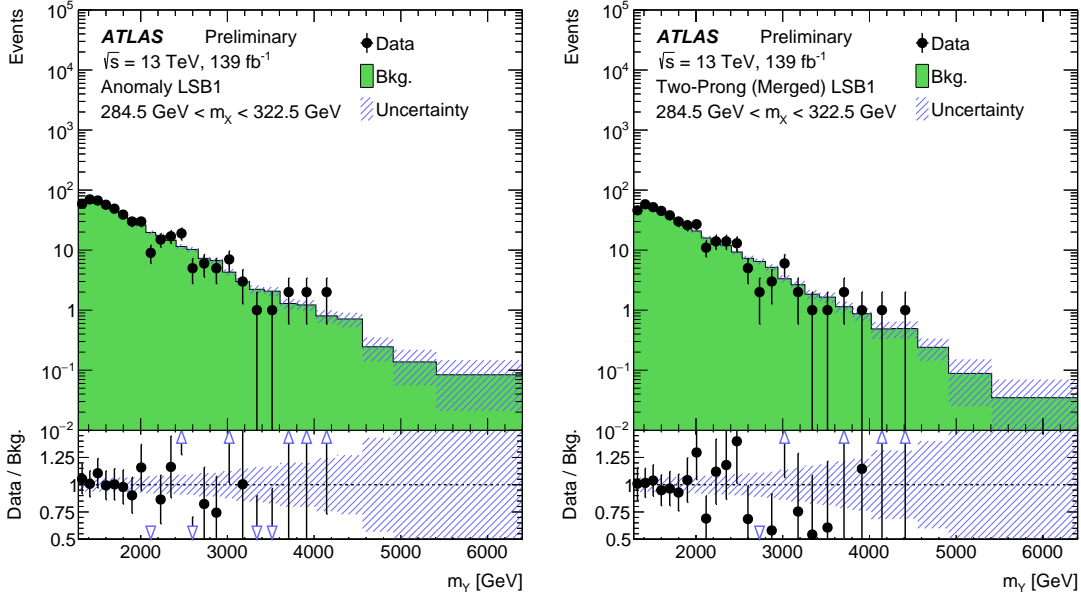
(a)

(b)
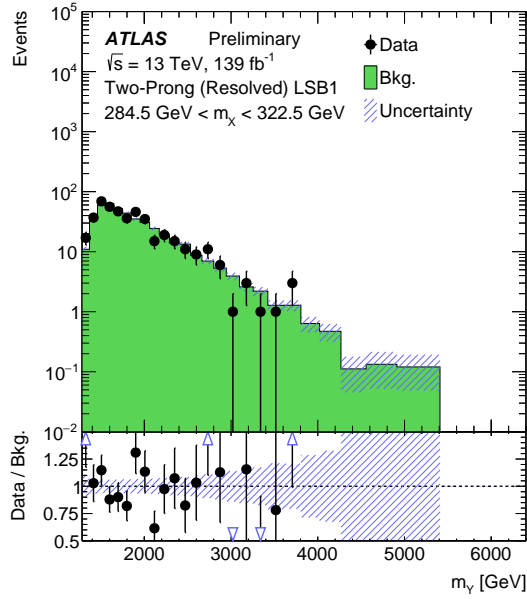
(c)

Figure 7: Reconstructed m_Y distributions of the background as determined by a background-only fit and data in the LSB1 VR, for the anomaly (a), two-prong merged (b), and two-prong resolved (c) selections, in the m_X bin [284.5, 322.5] GeV. The ratio of the observed data to the background is shown in the lower panel. The uncertainty band includes both statistical and systematic effects.
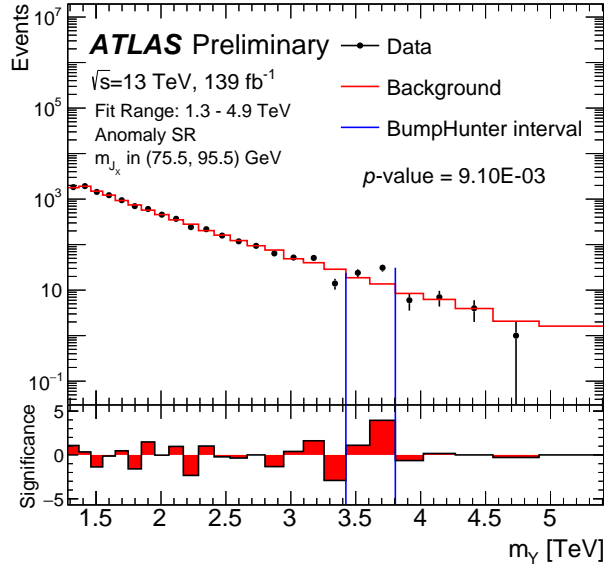
Figure 8: The m_{JJ} distribution associated to the m_X bin [75.5, 95.5] GeV in the anomaly SR, where the background is determined by a background-only fit to data with both statistical and systematic uncertainties included. The p-value of 9.1×10^{-3} pertains to the interval of m_Y bins chosen by `BumpHunter`, shown in blue, to have the most significant excess over expected background. The p-value and the lower panel of per-bin significances are computed with statistical uncertainties only. Incorporating the trials factor for all m_Y and overlapping m_X bins in the search, the single bin with the most significant excess has a global significance of 1.47σ.
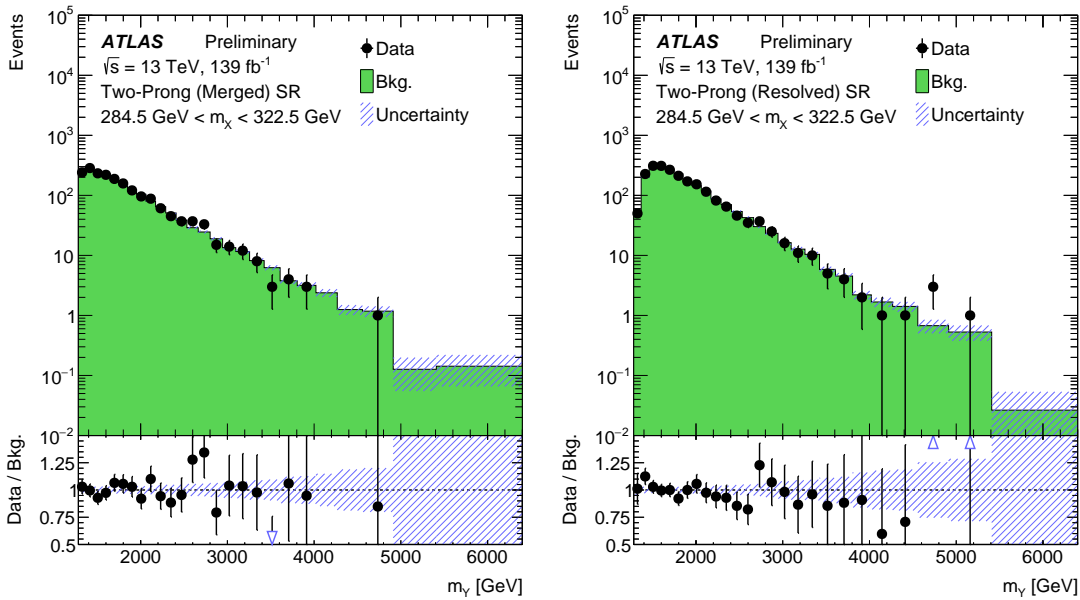
Figure 9: Reconstructed m_Y distributions of background and data in the SR, for the merged (left) and resolved (right) selections, in the m_X bin [284.5, 322.5] GeV. The background is determined by a background-only fit to the data with all statistical and systematic uncertainties included. The ratio of the observed data to the background is shown in the lower panel. The uncertainty band includes both statistical and systematic effects.
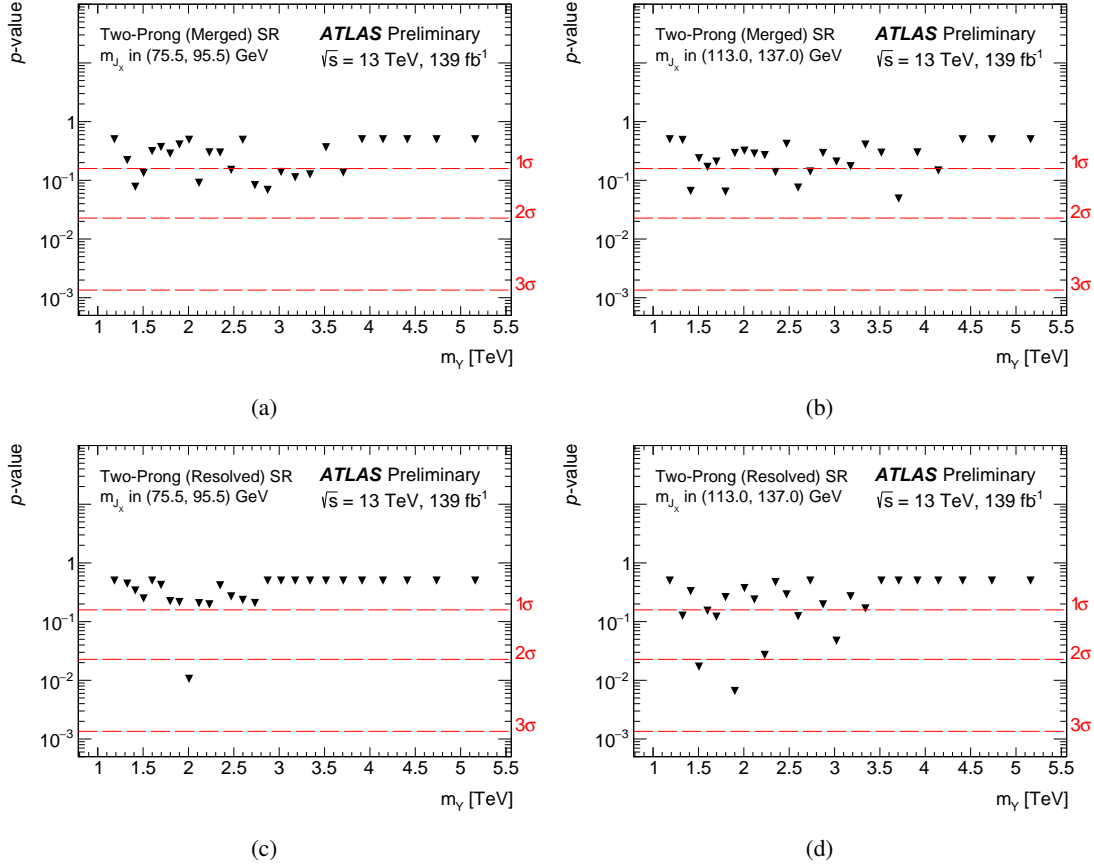
(a)

(b)

(c)

(d)

Figure 10: The p-value per m_Y bin for both two-prong SRs, calculated using only statistical uncertainties. 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 are thus split into merged W/Z window (a), merged Higgs window (b), resolved W/Z window (c), and resolved Higgs window (d). 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 approximates a constant value of 0.5 for the high Y mass region of the resolved SR, as this region of phase space is far more likely to produce a highly boosted J_X that falls in the merged SR selection.
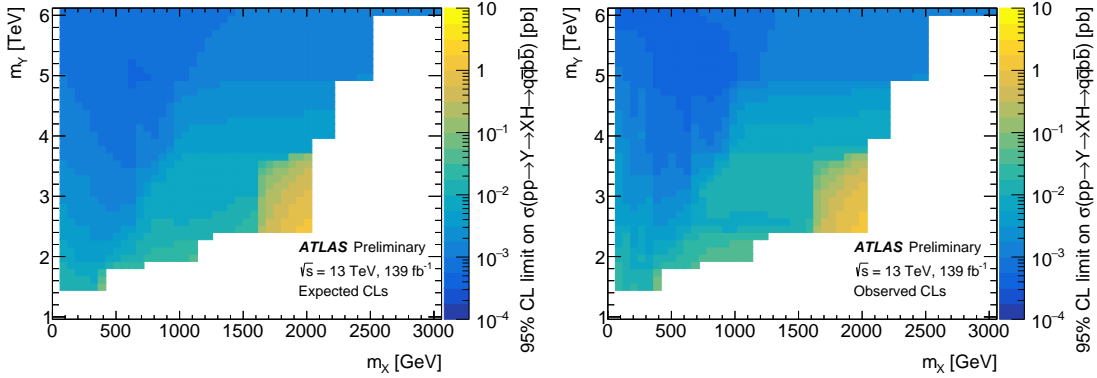
Figure 11: The expected (left) and observed (right) 95% CL limits on the cross-section $\sigma(pp \to Y \to XH \to q\bar{q}b\bar{b})$ in pb in the two-dimensional space of m_Y versus m_X, obtained from a simultaneous fit of both merged and resolved two-prong signal regions with all statistical and systematic uncertainties. A bilinear interpolation procedure is applied to provide results in between fully simulated signal points. The observed limits range from 0.342 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).

final states, only the merged two-prong region is considered for this comparison. 95% CL upper limits on the production cross section of six benchmark signals, including three $Y \to XH$ points and the three alternate jet topologies, are generated for both of these SRs. As systematic uncertainties on the signal efficiency of the anomaly score are not assessed, this comparison is performed using only statistical uncertainties and a post-fit background estimation in the limit calculation. Since the merged region uses D_2^{trk} and thus explicitly tags on the two-prong substructure of the X in the generated $Y \to XH$ grid, it is possible that these regions will outperform the fully unsupervised approach on the $Y \to XH$ signals. For points where the X is highly boosted and thus the anomaly score is most sensitive, the upper limit on the cross section is approximately the same across the merged and anomaly SRs, while the two-prong D_2^{trk} selections together give stronger limits across the rest of the $Y \to XH$ generated phase space. The signal model-independent aspect of the anomaly detection used in the anomaly SR is evident through improved limits on the alternative substructure signals. Notably, for the dark jets this factor of improvement is approximately 20, which underlines the strength of the model-independent approach particularly for signals that are challenging to characterize with existing high-level variables.

While these results cover areas of phase space that have not been previously studied directly by other searches, some analysis selections are highly correlated to those of other recent ATLAS dijet resonance searches. The m_X bin of [75.5, 95.5] would be sensitive to the VH resonance hadronic final state, which is covered by a dedicated analysis using the same dataset [57]. The approach here differs in both vector boson tagging and Higgs boson tagging approaches, 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 [58], though a direct comparison is not provided due to the assumptions made here of the Higgs boson mass and decay.

9 Conclusion

A search is performed for a heavy new boson Y decaying to a new particle X and a Standard Model Higgs boson in 139 fb^{-1} of LHC Run 2 data collected by the ATLAS detector. The analysis focuses on a fully

hadronic final state, where the *X* 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 over jets in data to define an anomaly detection SR, which selects the *X* particle solely based 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 are observed in the data with respect to the predicted background. The largest excess is found in the anomaly SR with a global significance of 1.47 σ considering all m_X and m_Y bins, and is not found to be compatible with the expected signal shape. Results are interpreted as upper limits at 95% confidence 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 lowest limit of 0.342 fb is achieved in the merged regime for the signal point (m_Y = 5000 GeV, m_X = 600 GeV), while the highest of 1.22 pb is achieved for the highly resolved point (m_Y = 2500 GeV, m_X = 2000 GeV).

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on p. 2).

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 2).

[3] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders*, Z. Phys. C **45** (1989) 109, [Erratum: Z.Phys.C 47, 676 (1990)] (cit. on p. 2).

[4] K. S. Agashe et al., *LHC signals from cascade decays of warped vector resonances*, Journal of High Energy Physics **2017** (2017), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP05(2017)078 (cit. on p. 2).

[5] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*, Physical Review Letters **83** (1999) 3370, ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.83.3370 (cit. on p. 2).

[6] T. Han, J. D. Lykken, and R.-J. Zhang, *Kaluza-Klein states from large extra dimensions*, Physical Review D **59** (1999), ISSN: 1089-4918, URL: http://dx.doi.org/10.1103/PhysRevD.59.105006 (cit. on p. 2).

[7] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics Reports **516** (2012) 1, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002 (cit. on p. 2).

[8] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, *Heavy vector triplets: bridging theory and data*, Journal of High Energy Physics **2014** (2014), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP09(2014)060 (cit. on pp. 2, 4).

[9] ATLAS Collaboration, *A search for resonances decaying into a Higgs boson and a new particle X in the XH→qqbb final state with the ATLAS detector*, Physics Letters B **779** (2018) 24, ISSN: 0370-2693, URL: http://dx.doi.org/10.1016/j.physletb.2018.01.042 (cit. on p. 3).

[10] CMS Collaboration, *Search for new heavy resonances decaying to WW, WZ, ZZ, WH, or ZH boson pairs in the all-jets final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, tech. rep., CERN, 2022, URL: https://cds.cern.ch/record/2803845 (cit. on p. 3).

[11] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 3).

[12] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 3), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[13] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 3).

[14] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 4).

[15] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on p. 4).

[16] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 4).

[17] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 4).

[18] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 4).

[19] R. D. Ball et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, Nucl. Phys. B **849** (2011) 296, arXiv: 1101.1300 [hep-ph] (cit. on p. 4).

[20] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep., CERN, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 4).

[21] GEANT4 Collaboration, S. Agostinelli, et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 4).

[22] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 4).

[23] J. de Blas, J. M. Lizana, and M. Pérez-Victoria, *Combining searches of Z′ and W′ bosons*, Journal of High Energy Physics **2013** (2013), URL: https://doi.org/10.1007%2Fjhep01%282013%29166 (cit. on p. 4).

[24] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 4).

[25] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 4).

[26] ATLAS Collaboration, *Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **09** (2019) 091, [Erratum: JHEP 06, 042 (2020)], arXiv: 1906.08589 [hep-ex] (cit. on pp. 4, 13).

[27] M. Park and M. Zhang, *Tagging a jet from a dark sector with Jet-substructures at colliders*, Phys. Rev. D **100** (2019) 115009, arXiv: 1712.09279 [hep-ph] (cit. on p. 5).

[28] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 5).

[29] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 5).

[30] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2275636 (cit. on p. 5).

[31] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv: 0912.1342 [hep-ph] (cit. on p. 5).

[32] ATLAS Collaboration, *Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS*, tech. rep., CERN, 2015,
URL: https://cds.cern.ch/record/2041461 (cit. on p. 5).

[33] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 5).

[34] ATLAS Collaboration,
Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector, tech. rep.,
CERN, 2015, URL: https://cds.cern.ch/record/2037702 (cit. on p. 5).

[35] ATLAS Collaboration, *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*, tech. rep., CERN, 2020,
URL: https://cds.cern.ch/record/2724739 (cit. on p. 6).

[36] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970,
arXiv: 1907.05120 [hep-ex] (cit. on p. 6).

[37] D. Krohn, J. Thaler, and L.-T. Wang, *Jets with variable R*,
Journal of High Energy Physics **2009** (2009) 059,
URL: https://doi.org/10.1088/1126-6708/2009/06/059 (cit. on p. 6).

[38] A. Kahn, J. Gonski, I. Ochoa, D. Williams, and G. Brooijmans,
Anomalous jet identification via sequence modeling, Journal of Instrumentation **16** (2021) P08012,
URL: https://doi.org/10.1088%2F1748-0221%2F16%2F08%2Fp08012 (cit. on p. 7).

[39] J. Chung et al., *A Recurrent Latent Variable Model for Sequential Data*, 2015,
URL: https://arxiv.org/abs/1506.02216 (cit. on p. 7).

[40] A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy correlation functions for jet substructure*,
Journal of High Energy Physics **2013** (2013),
URL: https://doi.org/10.1007%2Fjhep06%282013%29108 (cit. on p. 7).

[41] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*,
Journal of High Energy Physics **2011** (2011), ISSN: 1029-8479,
URL: http://dx.doi.org/10.1007/JHEP03(2011)015 (cit. on p. 7).

[42] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets*, Lecture Notes in Physics (2019),
ISSN: 1616-6361, URL: http://dx.doi.org/10.1007/978-3-030-15709-8 (cit. on p. 7).

[43] S. Kullback, *Information Theory and Statistics*, New York: Wiley, 1959 (cit. on p. 8).

[44] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, 2013,
URL: https://arxiv.org/abs/1312.6114 (cit. on p. 8).

[45] T. Kanamori, S. Hido, and M. Sugiyama, *A least-squares approach to direct importance estimation*,
The Journal of Machine Learning Research **10** (2009) 1391 (cit. on p. 11).

[46] U. Michelucci and F. Venturini, *Estimating neural network's performance with bootstrap: A tutorial*,
Machine Learning and Knowledge Extraction **3** (2021) 357 (cit. on pp. 11, 13).

[47] F. Chollet et al., *Keras*, 2015, URL: https://keras.io (cit. on p. 11).

[48] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (2017),
arXiv: 1412.6980 [cs.LG] (cit. on p. 11).

[49] e. a. Martín Abadi, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/ (cit. on p. 11).

[50] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2677054 (cit. on p. 13).

[51] ATLAS Collaboration, *Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, tech. rep., CERN, 2021, URL: https://cds.cern.ch/record/2777811 (cit. on p. 13).

[52] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton–proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, JHEP **09** (2013) 076, arXiv: 1306.4945 [hep-ex] (cit. on p. 13).

[53] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex] (cit. on p. 14).

[54] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 14).

[55] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter*, 2011, URL: https://arxiv.org/abs/1101.0390 (cit. on p. 14).

[56] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693 (cit. on p. 14).

[57] ATLAS Collaboration, *Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **102** (2020) 112008, arXiv: 2007.05293 [hep-ex] (cit. on p. 19).

[58] ATLAS Collaboration, *Dijet Resonance Search with Weak Supervision Using $\sqrt{s}$ = 13 TeV pp collisions in the ATLAS detector*, Phys. Rev. Lett. **125** (2020) 131801, arXiv: 2005.02983 [hep-ex] (cit. on p. 19).