IL NUOVO CIMENTO **47 C** (2024) 113
DOI 10.1393/ncc/i2024-24113-1

Colloquia: IFAE 2023

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(*)

A. D'Avanzo on behalf of the ATLAS Collaboration
INFN Napoli and Università di Napoli, Dipartimento di Fisica - Napoli, Italy

received 13 February 2024

Summary. — A search is presented for heavy resonances decaying into a Higgs boson (H) and a new particle (X) in a fully hadronic final state with the ATLAS detector at the CERN Large Hadron Collider (LHC) (ATLAS Collaboration, *JINST*, **3** (2008) S08003). The full Run II of LHC is analyzed, corresponding to an integrated luminosity of 139 fb^{-1}. A novel discovery signal region is implemented based on a jet-level anomaly score for signal model-independent tagging of the boosted X boson, representing the first application of fully unsupervised machine learning to an ATLAS analysis. No significant deviation from the SM is observed, so the results are interpreted in upper limits at 95% of confidence level (C.L.) on the production cross section $\sigma(pp \to Y \to XH)$ (ATLAS Collaboration, *Phys. Rev. D*, **108** (2023) 052009).

1. – Motivation and kinematic signature of the search

The discovery of the Higgs boson in 2012 at the LHC [3,4] completed the sequence of particles predicted by the Standard Model (SM), but many open questions still suggest the need for a physics Beyond the Standard Model (BSM).

In this article, a search for new particles Y and X in the resonant process $Y \to XH \to q\bar{q}b\bar{b}$ is described, using data collected by the ATLAS detector during the full Run II (2015-2018) of the LHC, amounting to an integrated luminosity of $139 fb^{-1}$. A representative Feynman diagram of the searched process is show in fig. 1, where it is natural to expect that these new heavy particles may have decays to a Higgs boson due to its strong coupling to heavier particles. One of the main aspects of this analysis is the model-independent signal region (SR) defined by the application of an Anomaly Detection (AD) algorithm. It consists of a fully unsupervised approach (new to an ATLAS analysis) for the tagging of the X candidate hadronic decay solely based on its incompatibility with the SM signature. The lack of evidence for new interactions and particles has motivated the execution of such a generic search to complement the existing rigorous, model-dependent analysis program.

(*) IFAE 2023 - "Poster" session

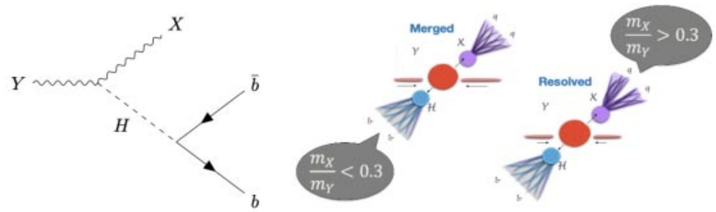

Fig. 1. – Feynman diagram of the target signal process (left). On the right, a schematic drawing of the two kinematic regimes in the signal model-dependent approach [2].

A model-dependent approach based on a Heavy Vector Triplet (HVT) [5] signal hypothesis is also considered, targeting the two-prong topology $X \to q\bar{q}$ in case no evidences are found in the anomaly signal region. The range of masses explored for the Y and X are respectively $1.5 < m_Y < 6$ TeV and $50 < m_X < 3000$ GeV, allowing for the X to be produced with a significant Lorentz boost if $m_X/m_Y < 0.3$. We so distinguish into a merged and resolved kinematic regime: one with collimated quarks so that the X can be reconstructed with a single large-radius jet and one with two small-radius jets from distant quarks respectively (fig. 1).

2. – Event selection and background estimation

The experimental signature $Y \to XH$ expects either two large-R jets with high p_T in the final state or one large-R jet and two small-R jets when $m_X/m_Y > 0.3$ in the model-dependent approach. This means that the Y resonance is reconstructed from such objects by considering the leading-p_T ones in each selected event. A preselection is also applied to assure the highest trigger efficiency for each collision event.

The unknown nature of the X particle requires that an assignment of the only known particle, the Higgs boson, is done before tagging the H and X candidates. This procedure is done by exploiting the outputs of a Deep Neural Network, trained to be sensitive to the jets substructure described by $H \to b\bar{b}$, that are used to define the discriminating variable D_{Hbb} [6]. The ambiguity between the two highest p_T jets is then resolved by considering the Higgs boson candidate as the one with the highest D_{Hbb} score, while the other is assigned as the X boson candidate. Once the Higgs boson candidate has been assigned, a cut on the Deep Neural Network output, $D_{Hbb} > 2.44$, defining a calibrated working point ([1]) with 60% efficiency is applied, enhancing the signal purity. (Figure 2).

The X candidate is tagged in three different ways based on the considered approach to the signal description. The first one is the anomaly detection, which targets the boosted decay of the X candidate through the prediction of a Variational Recurrent Neural Network (VRNN). This network is trained in a fully unsupervised way over jets with $p_T > 1.2$ TeV and modeled as a sequence of constituent four-vectors. A sensitive prediction, *i.e.*, the anomaly score (AS), to alternative X decay hypothesis other than two-prong (*e.g.*, heavy flavour, three prong, dark jet) is so achieved (fig. 2). The selection for such X tagging is defined with a flat cut on the X candidate $AS > 0.5$.

([1]) The calibration of the various flavour-tagging algorithms in ATLAS is usually performed only for a certain set of working points, each defined by a certain cut on the algorithm output score, corresponding to a certain tagging efficiency.

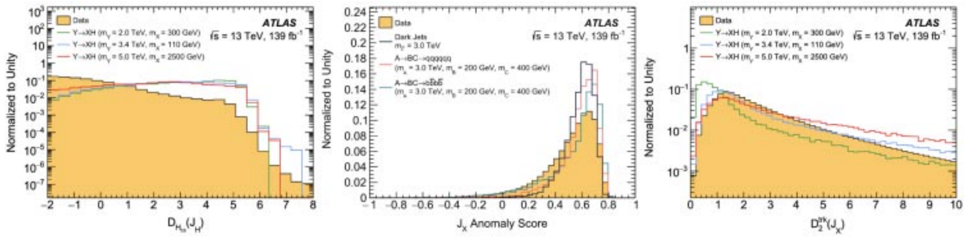

Fig. 2. – Distributions of final state jets variables for the three exploited X taggings, compared between data and several $Y \to XH$ signal points and exotic processes for the anomaly score output of the VRNN [2].

The second and the third are based on the identification of the X candidate two-prong substructure in the model-dependent merged and resolved regimes. It is done by means of the output of a multivariational algorithm on reconstructed jets which represents the likelihood of a jet to be composed by two collimated small-R jets. We call this output D^2_{Tracks} [7]. Lower values correspond to a higher likelihood that the X candidate is a large-R jet, meaning that it is perfectly suitable to select signal-like events in the merged regime. The distribution of the D^2_{Tracks} is shown in fig. 2, and the selection is applied with a flat cut on $D^2_{Tracks} < 1.2$. For signal points where the X can't be efficiently reconstructed with a large-R jet, a new assignment of the two small-R jets to the X candidate is performed. Then, the resolved selection is defined as orthogonal to the previous merged selection, with a flat cut at $D^2_{Tracks} > 1.2$. Additional cuts to improve the signal-to-background discrimination are applied, *i.e.*, the absolute value of the difference in rapidity $|\Delta Y| < 2.5$ and the p_T *balance* between the two small-R jets < 0.8.

Six analysis regions are finally defined based on the Higgs boson candidate mass and its b-tagging, as in fig. 3: the signal region, tagging events with the Higgs boson candidate mass in the Higgs boson mass window [75, 145] GeV and with a tagged Higgs boson candidate ($D_{Hbb} > 2.44$), and 5 ortogonal control regions, which are used for the background estimation procedure. The application of each X tagging selection means that overall each region splits into three others, increasing the total number to 18.

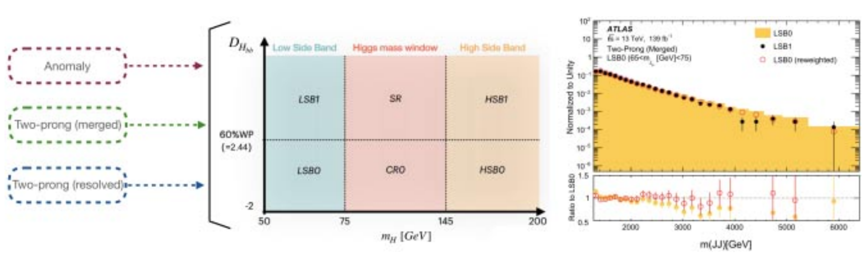

Fig. 3. – On the left, an illustration of the analysis regions for signal enhancing and background estimation also defined for every X tagging procedure used in this analysis. The m_{JJ} distribution in the LSB on the right shows the comparison between the shape of data before applying the reweighting function $w(x)$ for background estimation and after, represented by the yellow and red points respectively in the ratio plot below [2].

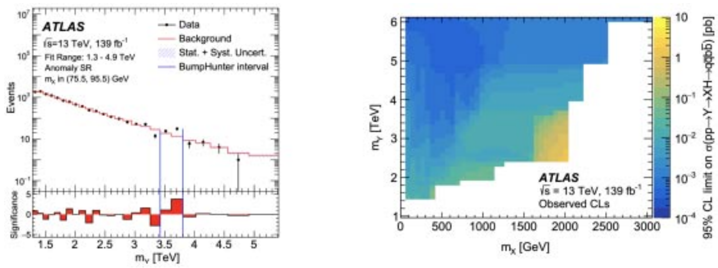

Fig. 4. – Distribution of m_{JJ} in the anomaly signal region for the m_X bin with the highest deviation from background (left); 95% C.L. upper limits on the production cross section for the process $pp \to Y \to XH \to q\bar{q}b\bar{b}$ across the signal grid (m_Y *vs.* m_X) (right) [2].

The background estimation in the SR is achieved with a totally innovative and fully data-driven technique, due to known mismodeling issues with simulated high-p_T multijet events. Distributions of data are reweighted from the control region 0 (CR0) to the SR with a function $w(x)$, learnt by a Deep Neural Network (DNN) in the training region between 145 and 175 GeV of the High Side Band (HSB), validated in the Low Side Band (LSB) between 65 and 75 GeV and finally extrapolated in the Higgs boson mass window. Generally good closure to data is observed in the validation region (LSB) for each X tagging scenario, considered inclusively to the network's training (fig. 3).

3. – Results

The results are extracted based on the final state di-jet invariant mass distribution m_{JJ} in the SRs as defined above, separately for each X candidate mass value. Two strategies are followed: model-independent discovery p-values are obtained by testing the background-only hypothesis in the anomaly SR, while model-dependent discovery p-values and exclusion limits are extracted through simultaneous profile-likelihood fits to the two two-prong SRs. Results are shown in fig. 4, where p-values in both approaches exclude significant deviations from the expected SM background, with a maximum deviation equal to 1.47 σ of global significance in the anomaly SR. For this reason, upper limits at 95% C.L. on the production cross section $\sigma(pp \to Y \to XH \to q\bar{q}b\bar{b})$ are determined for the whole 2D m_Y *vs.* m_X HVT signal grid considered (fig. 4).

REFERENCES

[1] ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[2] ATLAS Collaboration, *Phys. Rev. D*, **108** (2023) 052009.
[3] ATLAS Collaboration, *Phys. Lett. B*, **716** (2012) 1.
[4] CMS Collaboration, *Phys. Lett. B*, **716** (2012) 30.
[5] Pappadopulo D., Thamm A., Torre R. and Wulzer A., *JHEP*, **09** (2014) 060.
[6] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying Into $b\bar{b}$ with neural networks and variable radius subjets in ATLAS*, Technical Report No. ATL- PHYS-PUB-2020-019 (CERN) 2020.
[7] Larkoski A. J., Salam G. P. and Thaler J., *JHEP*, **06** (2013) 108.