PROCEEDINGS OF SCIENCE

Background estimation techniques in searches for heavy resonances at CMS

L. Benato* **on behalf of the CMS Collaboration**
Università degli Studi di Padova & INFN, Italy
E-mail: lisa.benato@cern.ch

Many Beyond Standard Model theories predict the existence of heavy resonances ($\geq$1 TeV) decaying into final states that include a high-energetic, boosted jet and charged leptons or neutrinos. In these very peculiar conditions, Monte Carlo predictions are not reliable enough to reproduce accurately the expected Standard Model background. A data-Monte Carlo hybrid approach (α-ratio method) has been successfully adopted since the LHC Run 1 in searches for heavy Higgs bosons performed by the CMS Collaboration. By taking advantage of data in signal-free control regions, determined exploiting the boosted jet substructure, predictions are extracted in the signal region. The α-ratio method and jet substructure techniques are described, along with some recent results obtained with 2016 Run 2 data collected by the CMS detector.

The European Physical Society Conference on High Energy Physics
5-12 July, 2017
Venice

*Speaker.

https://pos.sissa.it/

PoS(EPS-HEP2017)681

1. Introduction

Beyond Standard Model (BSM) theories try to overcome the problems left open by the SM (*i.e.*, the hierarchy of the Higgs boson mass) by enlarging the SM symmetry group. This will experimentally result in the appearance of new heavy resonances.

The Heavy Vector Triplet (HVT) model [1] is a framework summarizing different BSM theories. By adding few parameters to a simplified effective Lagrangian, it introduces a triplet of charged and neutral vector bosons (X^0, X^+, X^-). In the benchmark HVT-A model, the triplet decays mainly in fermions; in the benchmark HVT-B scenario, the dominant decays are into couple of vector and scalar bosons.

In this document, a background estimation technique suitable for searches for heavy diboson resonances is presented. The method is applied, in particular, to final states where a vector boson V, either a W or a Z boson, decays hadronically, and a Z boson decays in charged leptons (electrons or muons) or in neutrinos. Since the new particles are heavy (with a mass over 1 TeV), their decay products have a large Lorentz boost, hence the couples of quarks or leptons originating from the decay of the bosons are expected to be collimated. The V hadronically decaying boson is reconstructed as a large-cone jet; the Z leptonic decay is reconstructed as a pair of close-by leptons (if $Z \to \ell\ell$) or as a large amount of missing transverse momentum $\vec{p}_T^{\,\text{miss}}$ (if $Z \to \nu\nu$). The mass of the large-cone jet, interpreted as the V candidate, is used to define the signal region and the sidebands, namely signal-depleted control regions; data sidebands are used for the background estimation, that is performed with an hybrid data-simulation approach (α-method [2]). Jet substructure techinques allow to discriminate the V boson from the dominant backgrounds and increase the sensitivity of the searches. The considered analyses where the α-method is applied have been performed with 2016 data provided by LHC proton-proton collisions and collected by CMS detector [3], corresponding to an integrated luminosity of $\mathscr{L} = 35.9$ fb^{-1} ($Z \to \nu\nu$ [4]) or $\mathscr{L} = 12.9$ fb^{-1} ($Z \to \ell\ell$ [5]).

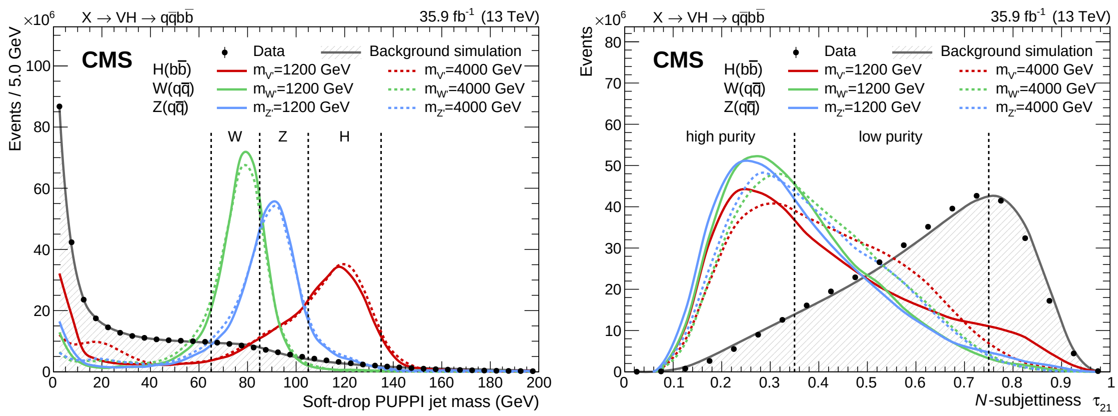

Figure 1: Variables used to tag an hadronically decaying boson in boosted topology, distinguishing signal events (coloured curves) and multijet background (in shaded grey). The jet mass (left [6]) defines the signal region of the analysis (W in green, Z in blue, H in red); the τ_{21} subjettiness (right [6]) is used to classify the events in exclusive categories.

PoS(EPS-HEP2017)681

2. Event reconstruction

A large-cone jet is clustered with the anti-k_T algorithm in a cone of 0.8, and is requested to have a transverse momentum $p_T > 200$ GeV. The boosted $Z \to \nu\nu$ candidate is reconstructed by asking $p_T^{\mathrm{miss}} > 200$ GeV, in a back-to-back topology with regards to the large-cone jet axis. Dedicated identification algorithms allow to efficiently reconstruct the couple of highly energetic leptons.

Grooming algorithms are applied to the jet. PUPPI algorithm [7] suppresses the contributions of particles coming from secondary events, not involved in the production of the heavy resonance. The softdrop algorithm [8] removes the soft radiation from the jet components. The groomed jet mass (m_j) is used to define the boson type: if $65 < m_j < 105$ GeV, the reconstructed jet is tagged as an electroweak boson (W or Z), if $105 < m_j < 135$ GeV it is tagged as an Higgs boson (fig. 1 [6], left).

The τ_{21} subjettiness variable [9] describes the sub-structure of the jet comparing the probability of being a 2-prong jet (typical of W and Z bosons) to the probability of being constitued by one monolithic structure (common feature of the multijet background events). The distribution of the τ_{21} subjettiness in genuine signal events is expected to peak at low values, while in the multijet background it peaks at higher value of the variable itself (fig. 1 [6], right). Events are classified in two exclusive categories: high purity category if $0 < \tau_{21} < 0.35$, dominated by signal events; low purity category when $0.35 < \tau_{21} < 0.75$. The categorization improves the discovery reach.

The analyses search for a local excess in data, in the invariant mass spectrum of the resonance candidate VZ, with regards to the predictions.

3. Background estimation technique and statistical analysis

The dominant background is represented by events where a leptonically decaying Z or W bosons are produced with jets ("V+jets"), and it is predicted with the α-method. Secondary backgrounds are represented by events containing a top quark ("Top"), and pairs of electroweak bosons ("VV"); their prediction relies on simulations.

Since the analyses aim at reconstructing the hadronic decay of an electroweak boson, the signal region (SR) is defined as $65 < m_j < 105$ GeV; sidebands (SB) are defined as $30 < m_j < 65$ GeV or $m_j > 135$ GeV. The α-method works in two steps.

1. The background jet mass distributions are fitted in simulations, with physical motivated functions, and added up together; the parameters of secondary backgrounds are fixed, whilst the parameters of the main background are extracted from a fit to data sidebands. The expected normalization in the SR is then predicted and compared with data in fig. 2 (top).

2. The background distributions of the invariant mass of the diboson candidate are fitted in simulations with exponentially falling functions. The α function is defined as the ratio of the shape of the V+jets background in the signal region and the V+jet background in the

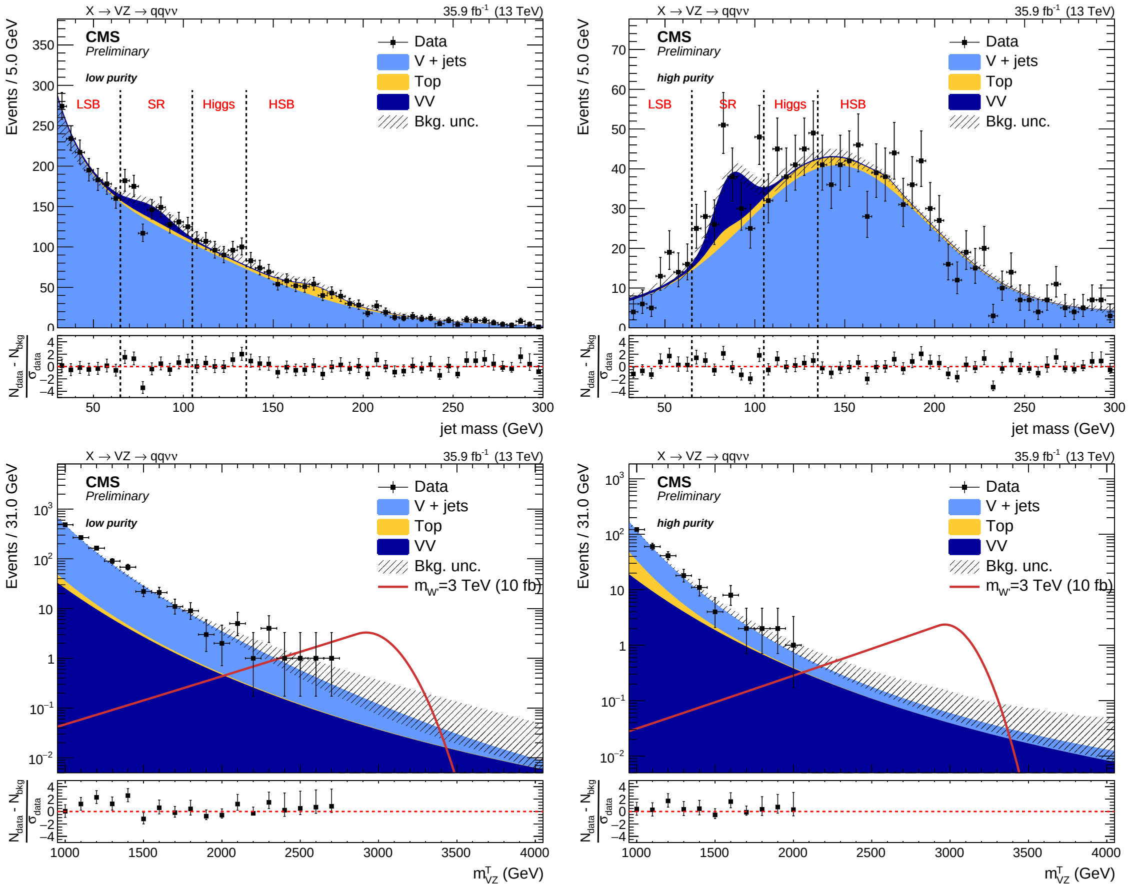

Figure 2: Top [4]: background distribution of the jet mass, representing the hadronically decaying V candidate, predicted with the α-method (histograms), compared to data (black markers). Bottom [4]: background distribution of the invariant mass of the di-boson candidate, predicted with the α-method (histograms), compared to data (black markers).

sidebands, calculated in simulations. The parameters of the secondary backgrounds are calculated in Monte Carlo simulations, while parameters describing the main background are extracted from a simultaneous fit to data SB and simulations. The expected background shape in SR is summarized in eq. 3.1, and compared to data in fig. 2 (bottom): α-ratio can be interpreted as a transfer function from SB to SR.

$$m_{VZ}{}_{SR}^{data} = \left[m_{VZ}{}_{SB}^{data} - m_{VZ}{}_{SB}^{Top,MC} - m_{VZ}{}_{SB}^{VV,MC} \right] \times \alpha(m_{VZ}) + m_{VZ}{}_{SR}^{Top,MC} + m_{X}{}_{SR}^{VV,MC}. \tag{3.1}$$

Uncertainties on the background predictions are extracted from the fit procedure on data (interpreted as statistical uncertainties), from the fit on simulations for the secondary backgrounds, and from the a-priori uncertainties on the physics objects involved (scales and resolutions of jets, mass of the jets, leptons, missing transverse momentum, τ_{21} data-simulations scale factors).

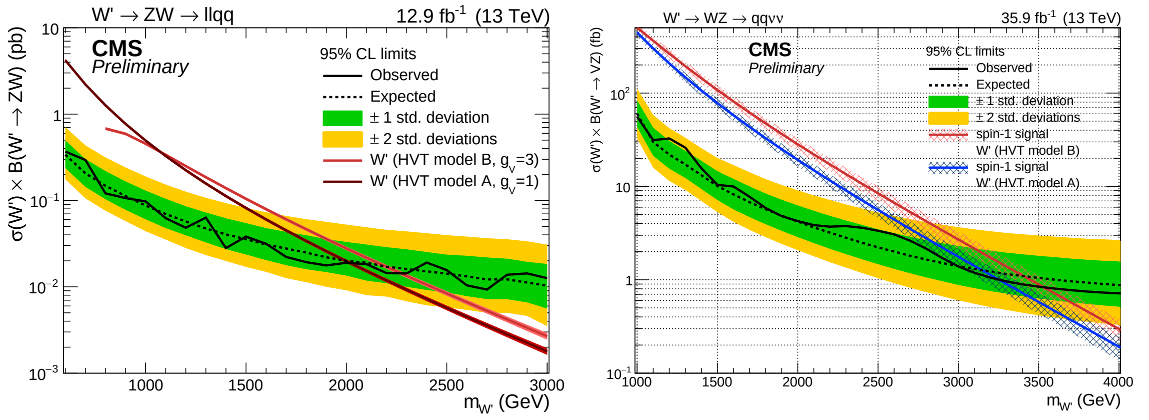

Figure 3: Expected (dotted line) and observed limits (solid line) on cross-section times branching fraction for a spin-1 W' predicted by the HVT model, as a function of the resonance mass (left: limits for the $VZ \to q\bar{q}\ell\ell$ final state [5], right: limits for the $VZ \to q\bar{q}\nu\nu$ final state [4]).

4. Results

No excesses are observed in data, with regards to the α-method predictions, neither in the $Z \to \ell\ell$ channel, nor in the $Z \to \nu\nu$ final states. Limits on the cross-section times branching fraction on spin-1 resonances expected by the HVT model are extracted with the asymptotic CL_S method; systematic uncertainties are treated as nuisance parameters. A W' of mass lower than 2.1 TeV (2.3 TeV), predicted by the HVT-A (HVT-B) scenario, is excluded by the analysis probing the $VZ \to q\bar{q}\ell\ell$ final state (fig. 3, left). A W' of mass lower than 3.1 TeV (3.5 TeV), predicted by the HVT-A (HVT-B) scenario, is excluded by the analysis probing the $VZ \to q\bar{q}\nu\nu$ final state (fig. 3, right).

References

[1] D. Pappadopulo, A. Thamm, R. Torre and A. Wulzer, *Heavy vector triplets: bridging theory and data*, *Journal of High Energy Physics* **2014** (2014) 1–50.

[2] CMS Collaboration, V. Khachatryan et al., *Search for a higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons*, *JHEP* **10** (2015) 144, [1504.00936].

[3] CMS Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[4] CMS Collaboration, *Search for heavy resonances decaying into a Z boson and a vector boson in the $\nu\nu$ $q\bar{q}$ final state*, CMS Physics Analysis Summary CMS-PAS-B2G-17-005, CERN, Geneva, 2017. http://cds.cern.ch/record/2273910.

[5] CMS Collaboration, *Search for heavy resonances decaying into a Z boson and a* W *boson in the $\ell^+\ell^-$*qq̄ *final state*, CMS Physics Analysis Summary CMS-PAS-B2G-16-022, CERN, Geneva, 2017. http://cds.cern.ch/record/2242955.

[6] CMS Collaboration, A. M. Sirunyan et al., *Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* **C77** (2017) 636, [1707.01303].

[7] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *Journal of High Energy Physics* **2014** (2014) 59.

[8] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [1402.2657].

[9] J. Thaler and K. Van Tilburg, *Identifying boosted objects with n-subjettiness*, *Journal of High Energy Physics* **2011** (2011) 15.