IL NUOVO CIMENTO **40 C** (2017) 87
DOI 10.1393/ncc/i2017-17087-4

COMMUNICATIONS: SIF Congress 2016

Search for heavy resonances decaying into a vector boson and a Higgs boson in the $(\ell\ell, \ell\nu, \nu\nu)$ $b\bar{b}$ final state

L. BENATO(1)(2) on behalf of the CMS COLLABORATION

(1) *Università degli Studi di Padova - Padova, Italy*
(2) *INFN, Sezione di Padova - Padova, Italy*

received 28 February 2017

Summary. — A search for heavy resonances in final states with a Higgs boson and a vector boson is presented, performed in the data collected with the CMS detector at $\sqrt{s} = 13\,\text{TeV}$ during 2015. The vector boson can be either a Z or W boson decaying leptonically (electrons, muons or neutrinos), whereas the Higgs boson decays hadronically into a couple of b-quarks of high momentum, detected as a single massive jet. The investigated final states consist of two b-quarks and zero, one or two charged leptons. Background shape and normalization are derived through a hybrid data-Monte Carlo method. The search is performed by examining the distribution of the reconstructed mass for a localized excess. Upper limits are derived as a function of the resonance mass and natural width, and are interpreted within the Heavy Vector Triplet theoretical model, as predicted in many scenarios beyond the standard model.

1. – Introduction

Many Beyond Standard Model (BSM) physics models have been formulated in the attempt of solving the puzzles left open by the standard model. A plethora of new heavy particles have been postulated, as a consequence of the enlargement of the standard model gauge symmetry group $SU(3)\times SU(2)\times U(1)$. The Heavy Vector Triplet (HVT) model [1] consists in a simple framework that generalizes a large number of theories, introducing a triplet of spin-1 bosons, that are heavier counterpart of the standard model $W^{\pm}$ and Z bosons. The model is parametrized by three parameters: g_V, describing the strenght of the interaction; c_H, the coupling with the standard model bosons; and c_F, the coupling with fermions. In the HVT model B, in particular, the coupling of the new resonances to bosons and fermions is comparable: $g_V = 3, c_H = 0.976, c_F = 1.024$. A significant fraction of these heavy vectors are expected to decay into a couple of bosons, namely the Higgs boson (H) and a vector boson (V, that can be either a W or a Z).

This analysis [2] searches for a heavy resonance decaying into an Higgs boson and a vector boson. The Higgs boson is reconstructed through its most probable decay channel,

i.e. a couple of b-quarks, whilst the vector boson is reconstructed via its leptonic decays, involving charged leptons (muons and electrons) and neutrinos, detected as missing transverse energy (E_T^{miss}). The probed final states are therefore three: $ZH \to \ell^+\ell^- b\bar{b}$; $W^\pm H \to \ell^\pm \nu b\bar{b}$; $ZH \to \nu\nu b\bar{b}$.

The search is performed by looking for a localized excess in the invariant mass spectrum (or transverse invariant mass) of the resonance candidate in data, compared to the standard model predictions. Data produced by LHC proton-proton collisions at a center-of-mass energy of 13 TeV have been collected by CMS detector in 2015, corresponding to an integrated luminosity of 2.2–2.5 fb^{-1}, depending on the probed final state. The main expected background contribution comes from the standard model production of a vector boson along with hadronic jets (V + jets). Since this background is poorly modelled by Monte Carlo simulations, a hybrid data-Monte Carlo approach (α-ratio) is adopted. It takes advantage of control regions in data to extrapolate the expected yield and shape of the V+jets background in the signal region. Top quark production represents the secondary background contribution, and it is predicted from simulation normalized to data in a proper top control region. Diboson processes involving vector bosons and Higgs boson (VV, VH) are a minor contributions, and they are evaluated with Monte Carlo simulations.

2. – CMS detector

CMS is a multi-purpose experiment [3] composed of many sub-systems enclosed in a superconducting magnet. Information coming from different sub-detectors is used together to improve the identification, energy and momentum determination of the particles involved.

3. – Event reconstruction

Dealing with heavy resonances means that the decay products of the vector bosons are expected to be highly boosted, hence close in angle. This leads to a non-trivial identification of the couple of quarks and leptons.

3˙1. *Higgs boson reconstruction.* – At CMS dectector, charged or neutral hadrons are clustered into jets using the anti-k_T algorithm [4]. Given the boosted topology, the Higgs boson is reconstructed as a large-cone jet, in which both the b-quarks are expected to lay. Jet energy corrections [5] are applied, in order to correct the jet 4-momentum by the non-uniform detector response as a function of the jet p_T and angular distribution, to remove the spurious energy deposits coming from secondary proton-proton interactions (pileup), and to correct the residual disagreement between data and Monte Carlo in the jet resolution.

The pruning algorithm [6] is applied, in order to remove contributions from soft radiation. The jet mass m_j is defined as the invariant mass of the pruned jet, and it determines the search region for the analysis, consisting into a jet mass range between 105 and 135 GeV in a final state exploiting an hadronically decaying Higgs boson. Two control regions, expected to be signal depleted, are defined: the low sideband, when $30 < m_j < 65$ GeV, and the high sideband, $m_j > 135$ GeV. Sidebands are used to predict the background normalization and shape (sect. **4**). The region $65 < m_j < 105$ GeV is kept blind, since it is the search region for other complementary diboson analyses, aiming at a final state with an hadronically decaying vector boson.

The combined secondary vertex algorithm [7] determines whether the large jet (or sub-jets) originates from the decay of a b-quark. Jet substructure techniques (softdrop algorithm [8, 9]) are utilized to re-cluster the jet into its constituents (sub-jets). The interplay between the b-tagging and the substructure of the Higgs jet allows to define the most effective approach to distinguish the background from a potential signal appearance in a wide invariant mass range of the diboson resonance. Two exclusive categories are defined, based on the number of b-tagged sub-jets (1 or 2). The 2 b-tagged sub-jets category is the most efficient at "low" masses, namely around 1–2 TeV, when the sub-components can still be distinguished in the large jet. When the resonance mass increases, the two b-quarks are strongly overlapped, therefore the softdrop and b-tagging algorithms are no longer able to find two different substructures in the jet. This approach guarantees an high signal efficiency at high masses (starting from 3 TeV).

Once reconstructed, the Higgs candidate is required to have a large boost ($p_T > 200\,\text{GeV}$).

3˙2. *Vector boson reconstruction.* – The V boson reconstruction depends on the number of charged leptons considered in the final state.

- 0ℓ: neutrinos do not leave any information in the detector, so they are reconstructed as missing transverse energy E_T^{miss}, defined as the magnitude of the negative sum of the transverse momenta of all the identified particles in the event, $\vec{p_T}^{\,\text{miss}} = -\Sigma_i \vec{p_T}^{\,i}$. Since, by construction, E_T^{miss} is strongly dependent on the other objects, a set of corrections is applied, in order to take into account the effects of the jet energy corrections, of the pileup, of the detector misalignment, of the unclustered energy deposits. Dedicated filters allows to reject events with large spurious missing energy due to detector noise and mismeasurement. When the Z decays into a pair of neutrinos, only the transverse momentum of the vector boson is available and it is identified as E_T^{miss}. Considering the boosted topology of the analysis, the p_T of the invisible Z should be larger than 200 GeV.

- 1ℓ: in a leptonic $W \to \ell\nu$ decay, a kinematical reconstruction is perfomed in order to extract the longitudinal component of the neutrino, p_z^ν. Once assumed that the neutrino x and y momentum components coincide with $E_{T,x}^{\text{miss}}$ and $E_{T,y}^{\text{miss}}$ respectively, p_z^ν can be calculated imposing the W mass constraint in the 4-momenta conservation equation, along with the condition $m_\nu = 0$: $m_W^2 = m_\ell^2 + m_\nu^2 + 2(E_\ell E_\nu - \vec{p_\ell}\cdot\vec{p_\nu})$. The real solution with the lower magnitude is chosen; else, if both the solutions are complex, only the real part is kept. The reconstructed W is required to have $p_T > 200\,\text{GeV}$.

- 2ℓ: the Z boson 4-momentum is reconstructed by adding the 4-momenta of the most energetic pair of opposite charge same flavour leptons. In order to improve the signal detection efficiency, electrons are required to satisfy loose identification criteria, and one of the muons can be reconstructed only in the inner silicon tracker. Leptons are required to be isolated (*i.e.* no energy deposits in a narrow cone around the lepton track); if they are overlapped in angle, the energy deposit of one lepton is subtracted from the other. The invariant mass of the pair of leptons should lay in a narrow window around the nominal mass of the Z boson, and the candidate should have $p_T > 200\,\text{GeV}$.

The diboson candidate is obtained by adding the 4-momenta (2ℓ or 1ℓ categories) or the p_T (0ℓ category) of its constituents, H and V. The variable in which the search is

performed is therefore the invariant mass of the VH candidate (2ℓ or 1ℓ categories), or the transverse mass (0ℓ category), defined as $m_T^{VH} = \sqrt{2E_T^V E_T^H \cdot (1 - \cos\Delta\phi(V,H))}$.

4. – Background prediction

Since the search aims at rare exotic phenomena, where small branching ratio decays are exploited, the expected event yield is small, therefore the background prediction plays the most important role in the analysis.

The α-ratio method takes advantage of jet mass sidebands in data to predict both the normalization and shape of the V + jets background. In the 0ℓ channel, the $Z \to \nu\nu$ and $W \to \ell\nu$ decays in association with jets represent the dominant backgrounds, whilst in the 2ℓ and 1ℓ channels the main contributions come from $Z \to \ell\ell$ and $W \to \ell\nu$ + jets. Given the common nature of the three processes (production of an electroweak boson decaying into leptons in association with a jet), they are considered toghether and referred to as V+jets or main background.

Single-top and $t\bar{t}$ processes bring sizeable contributions especially in the 0ℓ and 1ℓ categories. The normalization of the top quark background rely both on data and Monte Carlo distributions. Dedicated top control regions are built by asking additional b-tagged jets in the event, outside the cone of the Higgs jet. Data are in agreement with Monte Carlo distributions in these top control regions, both in the m_j and in the m_{VH} spectra. Scale factors are extracted to correct the Monte Carlo top event yield predictions.

Diboson events (labelled as VV, but including Higgs boson production as well) have a smaller impact. They are completely predicted by simulations.

4˙1. *Main background normalization.* – The background normalization is evaluated in the jet mass spectra. A fit to m_j is performed in Monte Carlo samples for the three background categories (V + jets, top and diboson), by using empirical probability density functions (p.d.f.s). Then, the m_j spectrum is fitted in data sidebands with the sum of the three p.d.f.s, in which the sub-dominant background parameters are fixed, whilst V + jets parameters are free to adjust according to data. The final predicted event yield in signal region is therefore

$$N_{SR}^{data} = \left[N_{SB}^{data} - N_{SB}^{Top} - N_{SB}^{VV}\right] + N_{SR}^{Top} + N_{SR}^{VV}\,.$$

The results of the fit to data are displayed in fig. 1, depending on the considered category ($0, 1, 2$ leptons; $1, 2$ b-tagged sub-jets). Background predictions in the signal region and data are in agreement.

4˙2. *Background shape.* – The sub-dominant background shapes are taken from simulations by fitting the invariant mass m_{VH} (or transverse mass m_{VH}^T) spectra with parametric p.d.f.s, generally modelled as smoothly falling exponential functions with high mass tails. The main background shape in the signal region is predicted by using a transfer function, the α-ratio, calculated on V + jets Monte Carlo samples, that accounts for the small kinematical discrepancies expected in the signal region with regards to the sidebands

$$\alpha(m_{VH}) = \frac{N_{SR}^{Vjet}(m_{VH})}{N_{SB}^{Vjet}(m_{VH})},$$

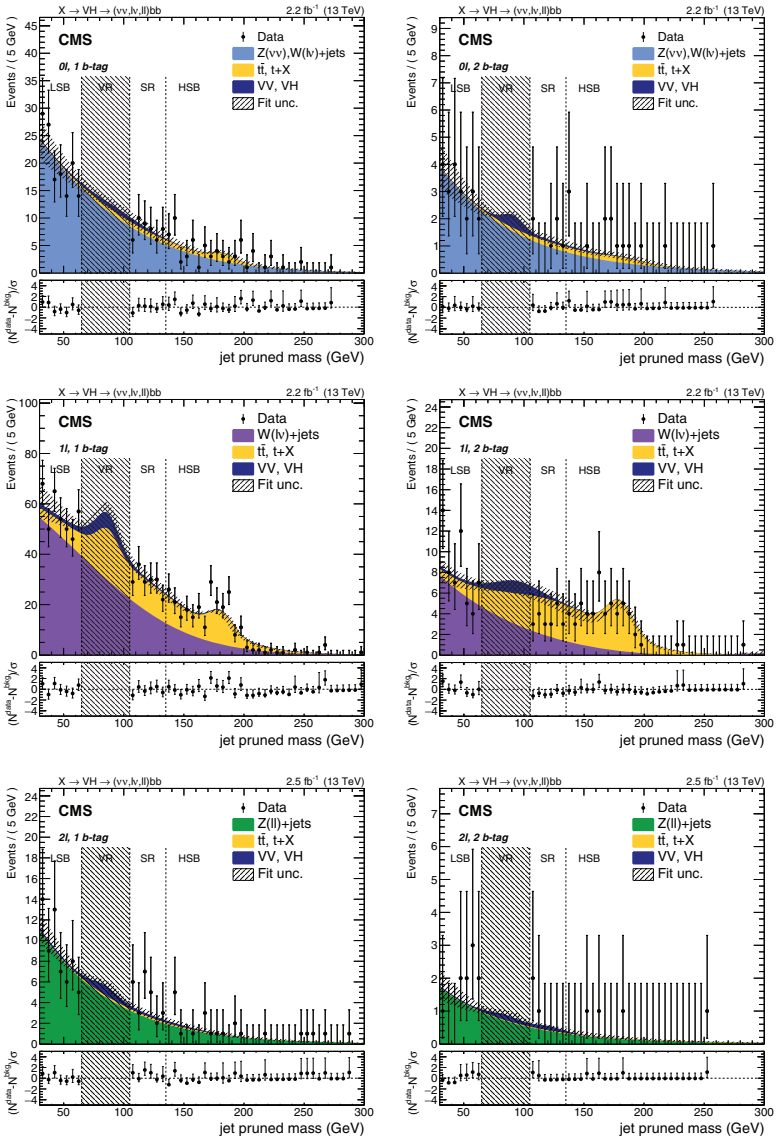

Fig. 1. – Jet mass distributions of the Higgs jet in the 0ℓ (upper), 1ℓ (middle), and 2ℓ (lower) categories, and separately for the 1 (left) and 2 (right) b-tagged sub-jet selections. The shaded band is the uncertainty from the fit to data sidebands. Black dots represent data, along with their Poisson uncertainties. Sidebands and signal region are separated by dashed vertical lines. The bottom pads displays the pulls distributions, namely $(N^{\rm data} - N^{\rm bkg})/\sigma$, where σ is the Poisson uncertainty in data.

Once defined the total background shape as the sum of the p.d.f.s of the three backgrounds,

$$N^{bkg}_{SR,SB}(m_{VH}) = N^{main}_{SR,SB}(m_{VH}) + N^{Top}_{SR,SB}(m_{VH}) + N^{VV}_{SR,SB}(m_{VH}),$$

a simultaneous fit is performed in data sidebands, in order to extract the V + jets background parameters, whilst the secondary background parameters are fixed. The final expected shape in signal region is then predicted by multiplying the fitted shape in data sidebands by the α-ratio

$$N^{data}_{SR}(m_{VH}) = \left[N^{data}_{SB}(m_{VH}) - N^{Top}_{SB}(m_{VH}) - N^{VV}_{SB}(m_{VH})\right] \times [\alpha(m_{VH})] + N^{Top}_{SR}(m_{VH}) + N^{VV}_{SR}(m_{VH}).$$

Results of the background shape in signal region are presented in fig. 2, depending on the considered category ($0, 1, 2$ leptons; $1, 2$ b-tagged sub-jets). No discrepancy is observed in data in the signal region, compared to the background predictions.

The robustness of the α-method is verified by splitting the lower m_j sideband in two sub-regions, and predicting shape and normalization of the intermediate sideband from the lower and upper sidebands. The distributions are in agreement.

5. – Systematic uncertainties

The statistical uncertainties due to the low event yield affect the sensitivity of the analysis more than the systematic uncertainties.

V + jets normalization uncertainties are dominated by the small number of events in data sidebands; a minor contribution comes from the propagation of the uncertainty due to the secondary background modelling. Top normalization is affected by the small number of events in top control regions. Diboson normalization is mostly affected by the theoretical uncertainty of the cross-section, and amounts to 20%.

V + jets shape uncertainties are taken from the covariance matrix of the fit to data sidebands.

The jet energy scale and resolution impacts both shape and normalization of signal and secondary backgrounds, and the effect is propagated to jet mass as well. The uncertainty amounts to a 5% variation in the event yield on background, 2–3% on signal, depending on the mass of the considered resonance. It affects the signal shape resolution by a 1%.

Uncertainties on the b-tagging efficiency are the largest source of normalization uncertainty for samples that are not normalized to data. For the signal efficiency, these uncertainties are between 4–15% and 8–30% in the 1 and 2 b-tagged sub-jet categories. For background events, they amount to 5 and 12%. An additional 10% b-tagging uncertainty is assigned to the $t\bar{t}$ background, related to the extrapolation from the top quark control region to the signal region.

The uncertainties in the trigger efficiency, lepton reconstruction, identification, and isolation amount to a 6–8% for 2ℓ-1ℓ categories, 3% in the 0 lepton category. In the former categories, the lepton energy scale and resolution are propagated to the signal shape, affecting the mean and the width of the signal model by a 16% and 10%.

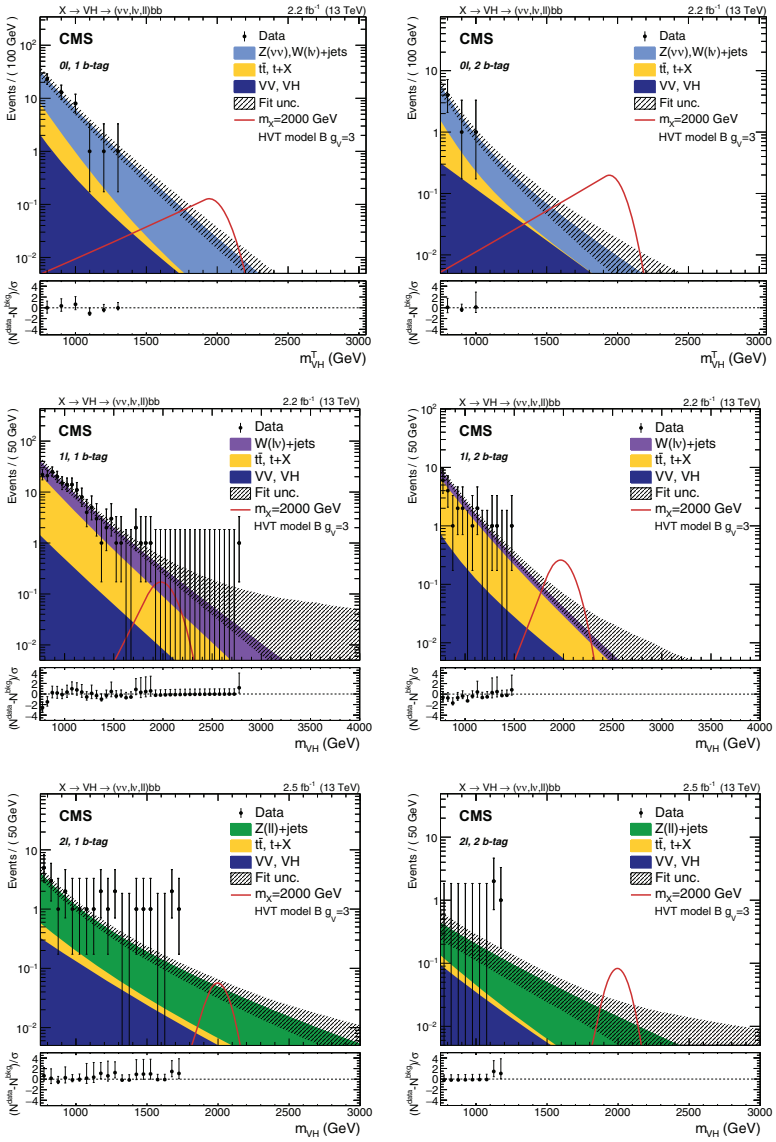

Fig. 2. – Resonance candidate mass distributions in the 0ℓ (upper), 1ℓ (middle), and 2ℓ (lower) categories, and separately for the 1 (left) and 2 (right) b-tagged sub-jet selections. The expected background events are shown with the filled area, and the shaded band represents the total background uncertainty from fit. Black dots represent data, along with their Poisson uncertainties. As a comparision, the expected distribution of a resonance with $m_X = 2000\,\text{GeV}$ produced in the context of the HVT model B is displayed with the solid curve. The bottom pads displays the pulls distributions, namely $(N^{\text{data}} - N^{\text{bkg}})/\sigma$, where σ is the Poisson uncertainty in data.

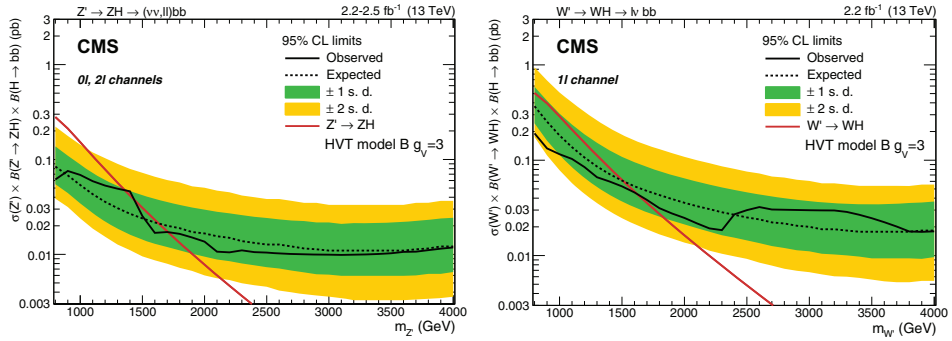

Fig. 3. – Observed and expected 95% CL upper limits on $\sigma(Z^{'}) \times BR(Z^{'} \to ZH) \times BR(H \to b\bar{b})$ (left) and $\sigma(W^{'}) \times BR(W^{'} \to WH) \times BR(H \to b\bar{b})$ (right) as a function of the resonance mass for a single narrow spin-1 resonance, including all statistical and systematic uncertainties. Bands represent the ± 1 and ± 2 standard deviation uncertainties on the expected limit. The solid curve corresponds to the cross-sections predicted by the HVT model B.

HERWIG [10, 11] parton showering algorithm has been applied alternatively to PYTHIA [12]. The relative difference is taken as a 7% uncertainty.

The factorization and renormalization scale uncertainties associated with the event generators leads to a 5% normalization variation in the diboson background, and a 4–12% on signal. The effect of these scale uncertainties is negligible on signal and background shapes.

Contributions on the normalization of backgrounds and signal come from pileup distribution uncertainty (3 and 0.5%), integrated luminosity (2.7%), E_T^{miss} scale and resolution (1% in 0ℓ), and the choice of parton distribution functions (3% for acceptance, and 4–18% for signal normalization).

6. – Results

The signal *versus* the background-only hypoteses have been tested through a profile likelihood fit, that has been performed to the unbinned invariant mass spectra of the reconstructed resonance. Signal is modelled as a crystal ball function, namely a Gaussian core with power law tails. Background shapes are predicted by the α-ratio method. Systematic uncertainties are treated as nuisances parameters and are profiled during the minimization. No significant excess has been observed in data, therefore a 95% confidence level upper limit is set on the signal HVT model B cross-section times branching ratio, by using the asymptotic modified frequentist method (CL_s) [13]. The observed and expected upper limits are displayed in fig. 3, for a spin-1 neutral $Z^{'}$ (obtained combining the 0ℓ and 2ℓ categories) and a spin-1 charged $W^{'}$ (1ℓ category). A more general result is pictured in fig. 4, where all the channels are combined together.

Exclusion limits can be set in the HVT parameters plane $[g_V c_H, g^2 c_F / g_V]$, where g is the electroweak coupling constant, and they are displayed in fig. 5. The grey region represents the parameter space where the width of the resonance is larger than the experimental resolution (5%), hence the narrow width approximation adopted in the signal modelling is no longer valid. The benchmark model B is shown as comparision.

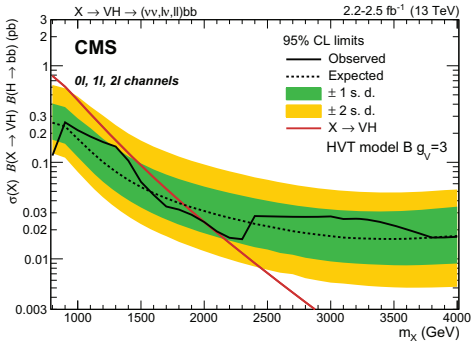

Fig. 4. – Observed and expected 95% CL upper limit with the ±1 and ±2 standard deviation uncertainty bands on $\sigma(X) \times BR(X \to VH) \times BR(H \to b\bar{b})$ in the HVT model B scenario, as a function of the resonance mass, for the combination of all the considered channels.

7. – Conclusions

A search for spin-1 heavy resonances decaying into a Higgs boson and a vector boson has been performed in a mass range of 0.8–4 TeV, using data produced by LHC proton-proton collisions at a center-of-mass energy of 13 TeV and collected with CMS detector in 2015, for a total integrated luminosity of 2.2–2.5 fb^{-1}. The exploited final states include leptonic decays of the vector boson ($\ell\ell$, $\ell\nu$, $\nu\nu$) and hadronic decay of the Higgs boson ($b\bar{b}$). Upper limits on the cross-section times branching ratio have been set in the range 10–200 fb; diboson resonances with masses lower than 2 TeV are excluded at 95% confidence level, in the context of the HVT model B scenario. Results restrict the parameter space available for new physics in the HVT framework, improving the reach of searches performed during LHC Run1 [14,15]. The final sensitivity is comparable with that obtained by a similar search performed by ATLAS Collaboration [16].

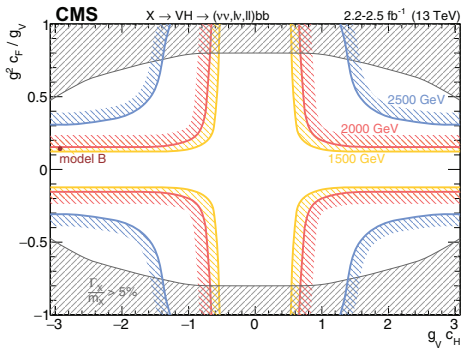

Fig. 5. – Observed exclusion in the HVT parameter plane $[g_V c_H, g^2 c_F/g_V]$ for three different resonance masses (1.5, 2.0, and 2.5 TeV). The benchmark scenario B with $g_V = 3$ is represented by the point. The gray shaded area corresponds to the region where the resonance natural width is larger than the experimental resolution (5%), therefore the narrow-width approximation fails.

REFERENCES

[1] Pappadopulo D., Thamm A., Torre R. and Wulzer A., *JHEP*, **09** (2014) 60.
[2] CMS Collaboration, *Phys. Lett. B*, **768** (2017) 137.
[3] CMS Collaboration, *JINST*, **3** (2008) S08004.
[4] Cacciari M., Salam G. P. and Soyez G., *JHEP*, **04** (2008) 063.
[5] CMS Collaboration, *JINST*, **12** (2017) P02014.
[6] Ellis S. D., Vermilion C. K. and Walsh J. R., *Phys. Rev. D*, **81** (2010) 094023.
[7] CMS Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2*, CMS-PAS-BTV-15-001 (2016).
[8] Dasgupta M., Fregoso A., Marzani S. and Salam G. P., *JHEP*, **09** (2013) 029.
[9] Larkoski A. J., Marzani S., Soyez G. and Thaler J., *JHEP*, **05** (2014) 146.
[10] Bellm J. *et al.*, *Eur. Phys. J. C*, **76** (2016) 196.
[11] Bähr M. *et al.*, *Eur. Phys. J. C*, **58** (2008) 639.
[12] Sjostrand T., Mrenna S. and Skands P., *Comput. Phys. Commun.*, **178** (2008) 852.
[13] Read A. L., *J. Phys. G*, **28** (2002) 2693.
[14] CMS Collaboration, *Eur. Phys. J. C*, **76** (2016) 1.
[15] CMS Collaboration, *JHEP*, **02** (2016) 145.
[16] ATLAS Collaboration, *Phys. Lett. B*, **765** (2017) 32.