PROCEEDINGS OF SCIENCE

Searches for BSM Higgs bosons in the VV, Vh, and hh final states in CMS

Roberto Covarelli*, **on behalf of the CMS collaboration**
Università and INFN Torino, Via Pietro Giuria 1, Torino, Italy
E-mail: roberto.covarelli@cern.ch

We present recent searches for high-mass scalar bosons (H) decaying to light-Higgs (h) or gauge (V) bosons in the CMS experiment, using 13 TeV proton-proton collision data from the RunII of the LHC machine. Namely, searches for $H \rightarrow ZZ$, $A \rightarrow Zh$, and $H \rightarrow hh$ are reported. Model-independent constraints on cross-sections or parameter exclusions in specific theory frameworks beyond the Standard Model, such as 2-Higgs-Doublet Models, are presented.

Prospects for Charged Higgs Discovery at Colliders - CHARGED2018
25-28 September 2018
Uppsala, Sweden

*Speaker.

PoS(CHARGED2018)023

1. Introduction and overview of theoretical models

High-mass scalar bosons (H, A) decaying to light-Higgs (h) or gauge (V) bosons are predicted in many extensions of the Standard Model (SM).

In the Minimal Super-Symmetric Model (MSSM), or its realization in terms of just Higgs-sector quantities (hMSSM) [1], the single-digit $\tan\beta$ region is best probed by such decay channels, as the decoupling limit is "delayed" to somehow large values of m_A and the $\tan\beta$ suppression is small. $H \to VV$ and $A \to Zh$ branching ratios (BR) can be still significant, while the width of these resonances remains negligible compared to typical experimental resolutions.

In 2-Higgs-Doublet Models (2HDM) the potential of diboson channels depends on the choice of model parameters (as shown in e.g. [2]) and is enhanced in some specific ones, such as fermiophobic H [3]. In both real and complex Scalar-Singlet models [4, 5] VV and hh heavy-Higgs decays are dominant.

Using 13 TeV proton-proton collision data from the RunII of the LHC, experimental searches for these new particles have been performed in the CMS experiment [6]. No significant excesses have been found and results are presented either in terms of cross-section limits or constraints in beyond-the-Standard-Model (BSM) parameters. Recent CMS results include seraches for: $H \to ZZ \to 4\ell/2\ell 2q/2\ell 2\nu$, $A \to Zh \to 2\ell 2b/2\nu 2b$, and $H \to hh \to 2b2\gamma/4b/2b2\ell 2\nu/2b2\tau$. There are many other CMS analyses targeting the same final states. However, they normally address very high-mass ($>$ 1 TeV) resonances, where fully-hadronic final states become the most sensitive thanks to the use of large-area jets. These are not usually relevant for BSM scalar searches, therefore are not reported here.

2. Search for $H \to ZZ$

The $H \to ZZ$ analysis [7] consists of a combined search for a heavy scalar resonance to ZZ in the 4ℓ, $2\ell 2q$, and $2\ell 2\nu$ final states. A model-independent study is performed, in a wide range of masses (m_H in [m_L, 3 TeV]) and widths (Γ_H/m_H in [0, 30%]), where m_L equals 160 GeV for the 4ℓ channel, 300 GeV for the $2\ell 2\nu$ channel and 550 GeV for the $2\ell 2q$ channel.

Cross-section limits are obtained from 2-dimensional (1-dimensional for $2\ell 2\nu$) likelihood functions, where the two analysis variables are the reconstructed ZZ mass m_{ZZ} (the transverse mass for $2\ell 2\nu$) and a kinematic discriminant based on matrix-element analytical descriptions of signal and background processes.

The assumed production modes are either gluon-gluon- (ggF) or vector-boson-fusion (VBF), with a free cross-section fraction f_{VBF}. The resonance kinematics is modelled from the POWHEG generator at the next-to-leading-order in QCD. The modeling of m_{ZZ} begins with a generator-level model that is a function of m_H and Γ_H. In case of large resonance width, the interference with SM Higgs[1] and continuum production in the relative mode is fully taken into account. This calculation is achieved through the *MELA* package, which uses matrix-element amplitudes from the MCFM and JHUGen MonteCarlo (MC) tools. The kinematic discriminant definition also employs matrix-element probabilities at generator level from the *MELA* package. Efficiency and resolution effects are then added from detector simulation.

[1] The H-h interference is also relevant because of the large off-shell tail in the ZZ decay mode.

For the 4ℓ final state, 4 leptons (electrons or muons) are selected with dedicated identification cuts. For at least 2 of them, high p_T is required. Selected events are split in 3 analysis categories: those with 2 VBF-tagging jets, those with reduced electron selection, and all other events ("untagged"). $2\ell 2\nu$ events are identified by two same-flavour leptons and a large missing p_T. b-tagged-jet and 3-lepton vetoes reject $t\bar{t}$ and WZ events respectively. Events are then divided into 3 categories: 0 jets, 1 jet and VBF-tagged. The $2\ell 2q$ final state is actually investigated in two separate analyses covering regions with large or small p_T of the hadronic Z: the "resolved" analysis, with two anti-k_T jets ($R = 0.4$) detected and the "merged" analysis, where a single anti-k_T jet ($R = 0.8$) containing all Z-decay products is reconstructed. For merged events, selections on the large-jet "pruned mass" (i.e. the invariant mass of constituents after removing soft/wide-angle components) and "τ-subjettiness" are applied. In this case, the kinematic discriminant is built using subjet directions. 3 analysis categories are also used here: VBF-tagged, tagged as $Z \to b\bar{b}$, and untagged. Fig. 1 (left) shows the data distribution in the $2\ell 2q$ final state together with signal MC and the superposition of background components.

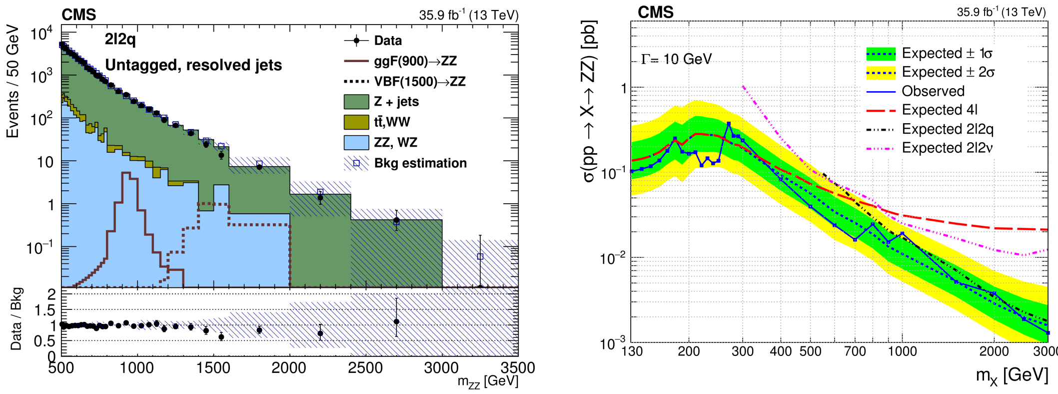

Figure 1: Left: Data distribution in the $2\ell 2q$ final state together with signal MC and the superposition of background components. Right: m_H projection of $(\sigma \times BR)$ 95%-CL limit for Γ_H = 10 GeV and with a final-state breakdown [7].

Background estimation is MC-based for most backgrounds (e.g. SM ZZ). For Z+jets control regions/samples in data are used, namely m_{jj} sidebands in $2\ell 2q$ and γ+jets in $2\ell 2\nu$. $(\sigma \times BR)$ 95%-CL limits are provided in the form of two-dimensional m_H-Γ_H plots from where constraints in a specific BSM model can be extracted. A m_H-projection example is provided in Fig. 1 (right) for Γ_H = 10 GeV and with a final-state breakdown.

3. Search for $A \to Zh$

The $A \to Zh$ analysis [8] is a search for high-mass resonances decaying to Zh in the $2\ell 2b$ or $2\nu 2b$ final states. It targets a specific BSM benchmark (2HDM) and the range of investigated masses is 225 GeV to 1 TeV. As a consequence of the model choice, both ggF and associated production with b quarks (bbA) is considered. Signal samples are generated with MadGraph/MadSpin at the leading order in QCD: the narrow-width assumption used is valid for $\tan\beta \gtrsim 1$ in the investigated m_A range.

The h signal region is defined to be in $100 < m(b\bar{b}) < 140$ GeV. Sidebands are used as control regions and a kinematic fit is applied to $m(Zh)$ constraining $m_h = 125$ GeV. In the $\nu\nu$ channel (used only in the high-mass region) the requirements $p_T(h) > 200$ GeV and $m_T(Zh) > 500$ GeV are applied. For the $\ell\ell$ channel, two likelihood-ratio MVA discriminators are defined, one based on angular variables (with a similar concept as *MELA*) and one based on $m(\ell\ell)$ and missing p_T (optimizing against top-quark background). 1-, 2-, and 3-b-tagged jets categories are defined in splitting events, the last targeting bbA production. Dominant backgrounds are Z+jets and $t\bar{t}$, whereas W+jets is only relevant in the $\nu\nu$ channel.

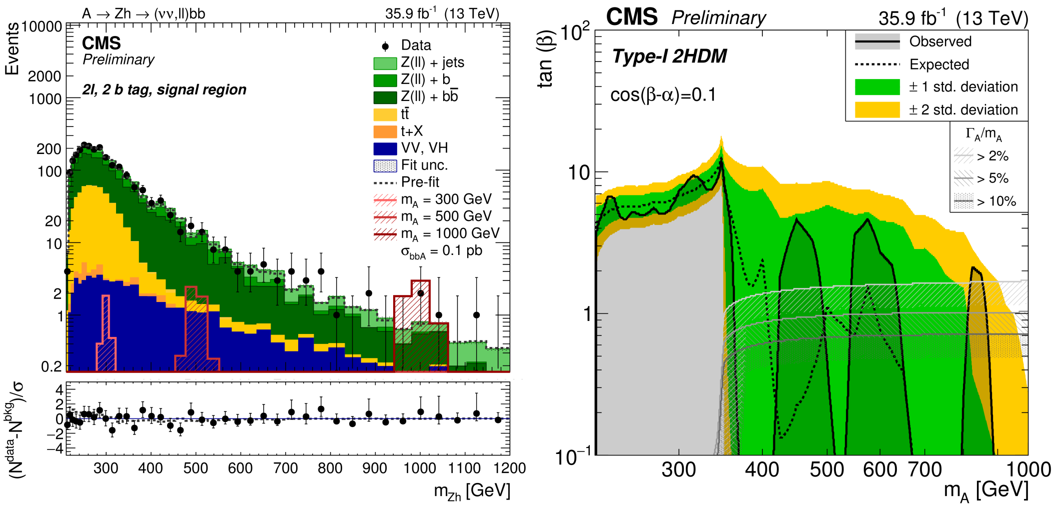

Figure 2: Left: $m(Zh)$ distribution in one of the analysis categories, the 2-b-tagged $2\ell 2b$. Right: Obtained limits in the $(m_A, \tan\beta)$ planes for type-I 2HDM [8].

By fitting simultaneously $m(Zh)$ distributions in signal region and yields in control regions, limits on $(\sigma \times BR)$ are obtained. Limits are then interpreted as constraints in the $(\cos(\beta-\alpha), \tan\beta)$ and $(m_A, \tan\beta)$ planes for all four types of 2HDMs. Figure 2 shows the $m(Zh)$ distribution in one of the analysis categories and the obtained limits in the $(m_A, \tan\beta)$ planes for type-I 2HDM.

4. Searches for $H \to hh$

There are several analyses in CMS addressing searches for $H \to hh$, depending on the hh final states: $2b2\gamma$ [9], $4b$ [10], $2b2\ell 2\nu$ [11], $2b2\tau$ [12], and their combination [13] are presented here. Other BSM scenarios with spin-2 resonances, or obtained from SM hh production by modifying the least experimentally-known h couplings, also contained in these works, are not relevant for scalar resonance searches.

The range of investigated masses is [$2m_h$, 900-1200 GeV] depending on the channel under study. Gluon-gluon-fusion $H \to hh$ is simulated with MadGraph at the leading order in QCD, with a narrow-width assumption (removing this assumption would require full calculation of interference with SM hh). The result is always given in terms of limits set on $(\sigma \times BR)$, which are compared to expectations for spin-0 radions in Warped Extra-Dimension models, but can be easily recasted.

The $H \to hh \to 2b2\gamma$ channel has a small yield because of the low $h \to 2\gamma$ BR, but excellent signal purity. The event selection is based on stringent b-tagging and photon identification/isolation criteria. Additional selection and categorization are performed by training a Boosted Decision Tree (BDT) whose variables are: jet b-tagging scores, p_T/m of $b\bar{b}$ and $\gamma\gamma$ pairs, and helicity angles. From the output of the BDT, high-purity and medium-purity categories are defined and used for signal search. The search is performed inside a region in the variable $\tilde{M} = m_{\gamma\gamma jj} - (m_{\gamma\gamma} - m_h) - (m_{jj} - m_h)$ whose limits are shifting to contain most of the signal for a given search mass. Signal extraction is performed through a 2-dimensional fit to $m_{\gamma\gamma}$ and m_{jj}. Fig. 3 (left) shows the data distribution in the high-mass/high-purity category together with the signal-background fit components.

The $H \to hh \to 4b$ channel has the largest BR, but huge QCD multijet background. The requirement of 4 resolved and b-tagged jets with invariant mass in a (large) region around 125 GeV is supplemented by the use of “multivariate regression” (i.e. dedicated jet corrections) and a double mass constraint on m_{h1}, m_{h2}. The signal region definition is a circle in the (m_{h1}, m_{h2}) plane: the part of the outer annulus where these two variables are less correlated is used as a sideband control region. Signal extraction is performed through a parametric fit to the m_{4b} distribution in 3 different mass ranges, after a thorough validation of background shape using sidebands. A small excess of data over the background fit prediction (around 460 GeV) has a 2.6σ significance.

The $H \to hh \to 2b2\tau$ channel has a BR which is intermediate, but features challenging final states for triggering and reconstruction. The analysis is performed in categories based on: τ final states ($e\nu\nu$, $\mu\nu\nu$, or hadronic), resolved or boosted b-jets and number of b-tagged jets (1 or 2). $t\bar{t}$ is the dominant background, suppressed using BDTs trained separately for each region, and is finally estimated from MC. The signal is extracted using distributions of the kinematically-fitted m_{hh}, where kinematic fits take into account missing momentum in the $\tau\tau$ system and double-mass constraints (in 2 different steps).

The $H \to hh \to 2b2\ell2\nu$ has a quite low BR and a challenging final state, having dileptonic $t\bar{t}$ as an irreducible background. Here the signal is extracted from a Deep Neural Network discriminator based on 8 variables and trained thorugh machine learning. The distribution of its output fitted in signal-enriched and 2 background-enriched $m_{b\bar{b}}$ regions. Backgrounds are estimated from MC except for a (small) contribution of Drell-Yan in the ee and $\mu\mu$ channels, computed from data control regions.

The combination of the expected and observed limits is shown in Fig. 3 (right): the first three decay modes contribute differently to the total limit in different regions of the phase space, while the $2b2\ell2\nu$ has somehow a smaller sensitivity.

References

[1] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGMSSMNeutral

[2] F. Kling, J.M. No and S. Su, *JHEP* 09 (2016) 093.

[3] D. Lòpez-Val et al., in arXiv:1610.07922.

[4] T. Robens and T. Stefaniak, *Eur. Phys. J* C76 (2016) 268.

[5] R. Costa et al., *JHEP* 06 (2016) 034.

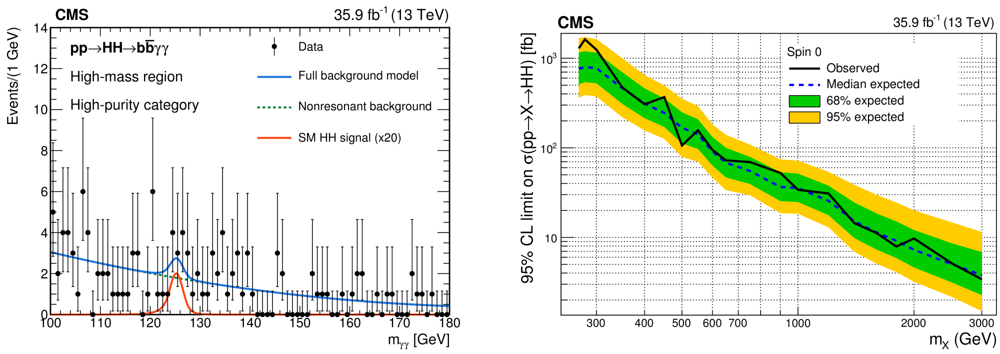

Figure 3: Left: Data distribution in the high-mass/high-purity category together with the signal-background fit components in $H \to hh \to 2b2\gamma$ [9]. Right: The combination of the expected and observed limits for the $H \to hh$ search [13].

[6] The CMS collaboration, *JINST* 3 (2008) S08004.

[7] The CMS collaboration, *JHEP* 06 (2018) 127.

[8] The CMS collaboration, CMS-PAS-HIG-18-005.

[9] The CMS collaboration, arXiv: 1806.00408, submitted to *Phys. Lett. B.*

[10] The CMS collaboration, *JHEP* 08 (2018) 152.

[11] The CMS collaboration, *JHEP* 01 (2018) 054.

[12] The CMS collaboration, *Phys. Lett. B* 778 (2018) 101.

[13] The CMS collaboration, CMS-PAS-HIG-17-030.